DOCUMENT RESUME

ED 025 858 EA 001 995

Evaluation and "Pace": A Study of Procedures and Effectiveness of Evaluation Sections in Approved Pace Projects, with Recommendations for Improvement. [Report No. 1 of the Second National Study of Pace.]

Center for Effecting Educational Change, Fairfax, Va.

Spons Agency-Office of Education (DHEW), Washington, D.C. Bureau of Elementary and Secondary Education.

Report No-ESEA-Title-3

Pub Date 29 Feb 68

Note-279p.

EDRS Price MF-$1.25 HC-$14.05

Descriptors-Computer Oriented Programs, Course Evaluation, *Creativity, Curriculum Development, *Educational Innovation, *Evaluation Needs, *Evaluation Techniques, Exceptional (Atypical), Information Dissemination, Information Needs, Inservice Teacher Education, Models, *Program Evaluation, School Community Relationship, Special Education, State Programs, Student Personnel Services, Supplementary Educational Centers

Identifiers-ESEA Title 3, PACE, *Projects to Advance Creativity in Education

This report studies the procedures and effectiveness of evaluation sections in projects approved under the ESEA Title III Projects to Advance Creativity in Education (PACE) program and makes recommendations for their improvement. Nineteen special consultants reviewed the evaluation sections of a random sample of 379 PACE proposals submitted in 1966 and 1967. Their ideas are classified into those dealing with (1) the need for and inadequacies of present evaluation procedures in the field of education, (2) evaluation of PACE projects, and (3) major problems found in PACE proposals. Without exception, the consultants criticized the effectiveness of PACE evaluation. Their analysis and criticism produced eight general recommendations that are designed to vitalize and improve the evaluation of PACE projects. These range from suggesting what proportion of project budgets should be devoted to evaluation to recommendations for establishing evaluation workshops. Finally, each of the consultants was asked to study 10 approved PACE proposals that focused upon his area of specialty, to judge the adequacy of evaluation procedures for these projects, and to prepare a report including any special recommendations for evaluation improvement. (TT)

ED025858

[illegible]

Approv [illegible]

Recommendati [illegible]

[illegible]

EA 001 095

EVALUATION AND "PACE":

A Study of Procedures and Effectiveness
of
Evaluation Sections
in
Approved PACE Projects,
with
Recommendations for Improvement

Richard I. Miller, Director

Special Consultants

William M. Alexander
Glenn O. Blough
Don Bushnell
Don Davies
Lloyd M. Dunn
Elliot W. Eisner
James D. Finn
Dorothy Fraser
Egon G. Guba
Robert J. Havighurst
Maurie Hillson
Arthur A. Hitchcock
Norman D. Kurland
John W. Letson
A. Harry Passow
Joseph B. Rubin
Wilbur Schramm
Ira J. Singer
Robert E. Stake

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

TABLE OF CONTENTS

Page No.

I. Forces Promoting Evaluation 1

Why the accelerating interest? 1
1. Increasing costs of education 2
2. Increasing complexity of education 2
3. Greater number of alternatives 3
4. Accelerating rate of obsolescence 3
5. Massive federal support 4
6. Greater concern about individualized learning 4
7. Greater use of academic findings from outside professional education 5
8. Greater assistance from outside professional education 5
9. Increased importance of education 6

II. Present Inadequaices in Evaluation 8

1. Difficulty of judging education 8
2. Fear of evaluation 10
3. Confusing action with accomplishment 11
4. Inadequate evaluative techniques 12
5. Inadequate teacher preparation 14
6. Shortage of qualified specialists 15

III. How Effective is PACE Evaluation? 19

Comparisons of 1966 and 1967 proposals 20

V. Major Problems in Proposals 23

Three major problems 23

1. Objectives 23
2. Procedures for evaluation 26
    A. Best procedures are not chosen 29
    B. Selected procedures are not closely related to activities 30
    C. Expert advice in developing procedures is not used 30
3. Amount budgeted 32

Solutions to meet the problems 33
A call for perspective 36

Page No.

V. Reports by Special Consultants

Curriculum Development--William M. Alexander A-1
Science Education--Glenn O. Blough A-13
Computer-Mediated Programs--Don D. Bushnell A-18
Inservice Teacher Education--Don Davies A-28
Special Education--Lloyd M. Dunn A-40
The Cultural Arts--Elliott W. Eisner A-55
Social Studies--Dorothy M. Fraser A-61
Urban, Metropolitan, and Rural Education Development--Robert J. Havighurst A-73
Organizational Flexibility--Maurie Hillson A-79
Pupil Personnel Services--Arthur A. Hitchcock....... A-89
Role of the States--Norman D. Kurland............... A-101
School-Community Relations--John W. Letson A-106
The Gifted and the Disadvantaged--A. Harry Passow .. A-129
Classroom Perspective--Joseph B. Rubin A-141
Communications--Wilbur Schramm A-147
Supplementary Service Centers--Ira J. Singer A-156
Institutionalization of Evaluation--James D. Finn A-170
An Evaluation Model--Robert E. Stake A-203
PACE Evaluation in General--Egon G. Guba A-219

INTRODUCTION

This is the first report of the second national study team of ESEA Title III. The study is sponsored by a Title III grant to the Center for Effecting Educational Change in Fairfax County, Virginia, which subcontracted the assignment to the University of Kentucky's Research Foundation.

It is fitting that the study team began its challenging work by focusing upon the difficult and critical problem of evaluation. As the following pages unfold, the reader may gain some insight into the task ahead of American education in this area. Our work is really a beginning; we are pleased with having grown individually and collectively but we are not pleased with the state of the art/skill of evaluation and we are not pleased with the small steps that we seem to have taken when giant strides are needed.

The first four chapters represent my efforts to snythesize and interpret the perceptive and challenging work by the 19 consultants in Chapter V. The list of the study team follows this introduction.

Richard I. Miller
Director

PACE NATIONAL STUDY

(1967-1968)

Special Consultants

William M. Alexander
Director, Institute for Curriculum Improvement
College of Education
University of Florida

Glenn O. Blough
Professor of Education
University of Maryland

Don Bushnell
Vice President and Director
Research and Development
Brooks Foundation
Santa Barbara, California

Don Davies
Executive Secretary, National Commission on Teacher Education and Professional Standards
National Education Association

Lloyd M. Dunn
Director, Institute on Mental Retardation and Intellectual Development
George Peabody College for Teachers

Elliot W. Eisner
Associate Professor of Education and Art
School of Education
Stanford University

James D. Finn
Chairman, Department of Instructional Technology
University of Southern California

Dorothy Fraser
Coordinator of Social Science
College of Education
Hunter College of the City University of New York

Egon G. Guba
Director, National Institute for the Study of Educational Change
Indiana University

Robert J. Havighurst
John Mosler Chair in Urban Education
School of Education
Fordham University

Maurie Hillson
Professor of Education
Graduate School of Education
Rutgers, The State University

Arthur A. Hitchcock
Professor of Education
State University of New York at Albany

Norman D. Kurland
Director, Center on Innovation
in Education
The State Education Department
Albany, New York

John W. Letson
Superintendent of School
Atlanta Public Schools

A. Harry Passow
Professor of Education
Teachers College, Columbia University
(on leave during 1967-68)
Department of Educational and
Psychological Research
School of Education
University of Stockholm

Joseph B. Rubin
Teacher, Chapman Elementary
School
Portland, Oregon

Wilbur Schramm
Director, Institute for
Communication Research
Stanford University

Ira J. Singer
Assistant Superintendent
West Hartford Public Schools

Robert E. Stake
Associate Director, Center
for Instructional Research
and Curriculum Evaluation
University of Illinois

Special Advisors

Harold B. Gores
President, Educational Facilities
Laboratories, Inc.
New York, New York

Donald W. Johnson
Director, Program Planning
and Development
California State Department of Education
Sacramento, California

Daniel L. Stufflebeam
Director, The Evaluation Center
The Ohio State University

PACE Project Directors' Advisory Group

Robert L. Hammond
Director, Project EPIC
Tucson, Arizona

Arthur R. King
Director, Hawaiian Curriculum Center
State Department of Education
Honolulu, Hawaii

Donald R. Miller
Director, Operation PEP
Burlingame, California

George G. Tankard, Jr.
Director, Center for Effecting Educational
Change
Fairfax, Virginia

Shelley Umans
Director, PACE Activities for
New York City
New York, New York

Study Director

Richard I. Miller
Director, Program on Educational
Change and Associate Professor
of Education
University of Kentucky

CHAPTER I

FORCES PROMOTING EVALUATION

Evaluation has become a number one concern for American education. It is very popular today, but one should be hesitant about placing it in the blossom-and-fade category of many other innovations and programs. Quite to the contrary, we are just at the edge of a surging interest in better evaluation both on part of educators and parents. As expressed by Ira Singer: "At long last, evaluation is 'in'. The school man has gotten the message. Beset by a bewildering array of how-to-do-it monographs, guides, models and flow charts, the practitioner has 'fixed' on this new star on the educational horizon. His initial peek is a furtive as a fugitive's look over the shoulder for the pursuing law. In a sense, he too has been getting away with murder for years, and knows it. The book is being thrown at him—in fact all the books from all sides and at one time." (Singer, A-156).

Why the Accelerating Interest?

There are several reasons for the current interest, and nine of these will be outlined in this section, but in a larger sense "this interest in evaluation," as expressed by James Finn, "must be seen in a technical-political context within the entire educational enterprise. The age of analysis in which we live is generating an age of assessment in education." (Finn, A-170).

1. Increasing Costs of Education

Education is becoming substantially more expensive at all levels, and an increasing number of communities are voting down tax increases—probably a higher percentage of vetoes than at any time in the past 20 years. Citizens have not adjusted comfortably to the increasing costs of education, which happens to be one of the few areas in which they can express their displeasure with taxes in general. They cannot do much against rising costs for living and pleasure but they can vote against school bond issues and other school tax efforts as general anti-taxation reactions.

Some school tax defeats are due to political factors and/or other circumstances entirely beyond the control of school officials such as general tax dissatisfaction, but many defeats are due to inept campaigns—and a major factor in determining the outcome of many campaigns is the quality of evaluative evidence. In other words, do school officials effectively present their case, including how past tax increases have served to improve education? Since this kind of evidence can only be obtained through evaluation; the use of this process becomes important for continued educational improvement.

2. Increasing Complexity of Education

Decisions about "best" courses of educational action are becoming more difficult as every dimension of living is more complicated. The rote lessons for learning spelling and arithmetic are being replaced by learning levels, by more individually prescribed instruction, by flexible scheduling, and the like.

These newer developments, however, bring greater need for appraisal as well as pose greater evaluative problems. Most good teachers and educators develop fairly reliable procedures for evaluating simple student learning but newer educational developments require considerably more sophisticated evaluation efforts. Also, procedures that would provide fairly reliable evaluative indexes some years ago are inadequate today, and they will be dangerously archaic tomorrow; therefore, carefully planned evaluative approaches become increasingly important.

3. Greater Number of Alternatives

Robert Frost's dilemma over "which path" presents a vastly more simple choice than the many options that now are open. In most areas of education— such as content selection, organization revision, and measurement— the array of choices requires greater evaluation as a basis for making intelligent decisions. This more traditional use of evaluation as judgment remains a vital dimension of evaluation.

4. Accelerating Rate of Obsolescence

Evaluation becomes more important as the length of useful life of educational innovations and materials decreases. To keep current with rapidly moving developments in science, for example, fundamental revisions in instructional materials should take place about every five years.

Accelerating rates of obsolescence call for earlier and more intensive procedures for evaluation. If there is much time lapse between the introduction of a new program and its evaluation, the introduction of still newer developments may not be realized or the change may not seem justified unless evaluation evidence is available. In brief, we need to know more—sooner—about new programs, thus the need for more attention to evaluation.

5. Massive Federal Support

The "pursuit of excellence" has been an educational battle cry since shortly following the Second World War, accentuated in the latter fifties by salvoes fired by Admiral Rickover and the launching of Sputnik. Unfortunately for central city youngsters, the quest has taken place several miles out—in the suburbs, or perhaps in medium sized cities that already had good schools.

The Elementary and Secondary Act of 1965 with almost one billion dollars per year in Title I for the educationally disadvantaged, the Office of Economic Opportunity (Poverty program) with about one billion dollars, and several other federal acts for vocational education focus upon lifting up the educationally disadvantaged. With massive new public monies in education, questions relating to cost-effectiveness and priorities are being asked with increasing frequency and persistence.

6. Greater Concern About Individualized Learning

Individualization, or adjusting instruction to the child, has received mountainous verbal support over the years, but only a few meaningful adjustments have been made to bring practice more in line with theory—until recently. Individualized prescribed instruction (IPI), continuous pupil progress (nongradedness), and programmed instruction are three examples, and each require more sophisticated and frequent evaluation than is the case with less complex programs. Evaluation becomes decision making as well as judgmental in these instances, serving to guide subsequent efforts.

7. Greater Use of Academic Findings From Outside Professional Education

The concept of evaluation as feedback and guide is borrowed from cybernetics, and the demand for varied types of evaluation comes from understanding of human behavior and learning. And computer technology has opened vast new possibilities for evaluation. National assessment, for example, was not possible before the advent of computer technology. Educational evaluation can be expected to benefit as advancements are made a vast array of fields, provided translators are "on the alert" for elements that might be transfered to education.

8. Greater Assistance From Outside Professional Education

Aerospace "brain" trusts are turning to education to sustain themselves largely because of the cutback in aerospace funds; and certain military branches are developing their own programs to bring achievement test scores of draftees above the minimum for military service. In addition, industry has moved into education in a major way, spending close to one billion dollars a year on educational hardware.[1] And several new educational consultant firms are featuring evaluation.

A number of ESEA Title I and III project directors have hired commercial firms to do their evaluation, believing that the outside expertise may be superior to that offered by professional educators or that educational evaluators are not available. More extensive use of "outside" elements can be expected, and it probably will bring new dynamism and creativity into the evaluation area.

9. Increased Importance of Education

Dating back to the "ole Satan deluder" Act of 1642, the American people have had a strong belief in the importance of education. One can find much fault in the intensity and consistency of this concern over the years, but the overall record must be considered quite admirable—especially in cross-cultural perspective.

The contemporary emphasis on educational importance can be traced to several factors. As mentioned earlier, the beep of Sputnik was a primary stimulus although considerable momentum for improvement was already evident by the Autumn of 1957. Focus on slum conditions in central cities; vast educational differences within any one state, usually greater than differences between states; increasing educational requirements for desirable employment; and increasing family mobility carrying with it demands for some equivalence in educational quality throughout the nation, or at least a higher minimum national quality—these factors are instrumental in prompting educators to evaluate their programs more vigorously than ever before. National assessment—an effort to provide some way of judging the quality of education on a national scale—definitely stems from these forces.

---

[1] John R. Stark, "Educational Technology: A Communications Problem," Phi Delta Kappan, 48:196; January 1967.

With cessation of hostilities in Vietnam, education may well move rapidly toward replacing national defense as the number one national expenditure. As education assumes this role, much greater attention will be directed toward assessing how well the monies are being spent.

These nine factors add up to an unprecedented concern about evaluation—about finding out how whether we are achieving what are set forth as goals or targets; if not, why not; and what corrections should be made to move the program back "on target." Those who believe the "evaluation itch" is another fad that will soon join educational history simply misread the times. Quite to the contrary, more likely we are standing on the threshhold of new understandings and developments in evaluation that can have significant bearing upon the course of our schools.

CHAPTER II

PRESENT INADEQUACIES IN EVALUATION

Why do we have a "sad face" about evaluation? Why has not the science and art of evaluation moved ahead more dynamically? This chapter will touch upon six reasons for the present inadequacies.

Difficulty of Judging Education

How can one ever be sure of evaluation where people are concerned? The only thing to fear is a presumption that we can touch all bases in evaluating people—in this case pupils and youngsters. We cannot now and probably will not be able in the foreseeable future to predict with finality or measure with precision human behavior and attitudes.

Some types of measurements, however, can be final and absolute. A football team either makes or misses a first down; a pianist plays the right or wrong note; and a word is spelled properly or improperly. But larger and less precise kinds of evaluations must be viewed with caution. For example, the following evaluation was made of the admissions application of Spencer Thompson:

> Latin teacher's report: "I have found the boy most difficult to teach. . . . He seems to have little or no understanding of the subject except in the most mechanical way. At times he seems almost perverse in his inability to learn. I suspect that he has received help from other boys in this prepared work."
>
> Headmaster's report: "Spencer is rather delicate owning to a severe pulmonary illness two years ago, but he seems to

> have recovered satisfactorily. He is too small to be effective in contact sports, but he greatly enjoys riding and swimming. The boy is certainly no scholar and has repeated his form twice. He does well in English, however, and possesses an excellent memory. In fact he won the School prize for reciting poetry last year. He has also, I regret to say, a stubborn streak, and is sometimes rebellious in minor matters, although he usually conforms. He is at once backward and precocious, reading books beyond his years, and yet ranking at the bottom of his form. . . He has, I believe, a native shrewdness and is a manly little fellow, high-spirited and well-liked, who unfortunately has not made the most of his opportunities here. I can recommend him to you on the grounds of general ability."

On the basis of this information and other data of a similar nature, Spencer Thompson was turned down by a private boy's school in New Hampshire. Only then did the Committee learn that Spencer Thompson was a pseudonym for a young schoolboy named Winston Spencer Churchill!

There are few Churchills or Edisons but how can we be sure that a similar pattern has not occurred countless times? It is good that we cannot relegate the human personality to a test tube, and educators need to be tough and uncompromising in limiting the presumptions of some evaluators who claim much more than they know.

This cautionary note, however, must not be interpreted as anti-evaluation— it is only against claiming too much for evaluation. The human element in evaluation *is* becoming more understandable as various technologies and techniques of evaluation move ahead. Significant strides *have* been made recently in mass sampling procedures aided by the computer and in new statistical techniques, and we know a great deal more about individualization of learning and the vast intra-differences within any one individual. This progress needs to be encouraged and assisted in every way possible.

Fear of Evaluation

While significant strides have been made in evaluation, some major problems remain, and one is simply the fear of evaluation. Arthur Hitchcock writes that evaluations are beset with three psychological fears:

> 1. Educators are unaccustomed to evaluation and the thought is scarey. A person in education is accustomed to action and interaction, and he can advance boldly on this front, and even advertise his action plans and procedures. But he feels uncomfortable with evaluation.
>
> 2. A certain defensiveness that evaluation cannot do justice to the project anyway.
>
> 3. A fear that the colossus. . . will conclude from the evaluation that the project is not as good as it really is. (Hitchcock, A-93)

Fear of evaluation is largely fear of the unknown, tinged with healthy skepticism for the rudimentary level of many evaluative procedures. Most public school officials are suspicious of "outsiders" who may be called in to evaluate. After all, local officials must live with the situation long after the "fireman" has caught flight 709. And the subtleties of a situation, which probably escape all but the most astute local officials, certainly remain unknown to the consultant.

These fears are real—as well as greatly inflated. The proportion of fear about evaluation usually is inversely proportion to knowledge about it. As one becomes knowledgeable about evaluation, its strengths and limitations become understood and evaluation becomes a procedure for providing feedback and guidance as well as judgment.

Confusing Action With Accomplishment

In this third annual address, President Millard Fillmore cautioned against mistaking change for progress.

A tendency does exist to confuse action with accomplishment or improvement, or to equate quantity with quality. If one is active enough, so this "logic" goes, then evaluation will somehow take care of itself. A PACE director who is working 10 to 12 hours a day, six days a week believes that "good" inevitably will evolve from such a dedicated effort. He also may believe that systematic procedures for evaluation will take valuable time and resources to learn what his staff already knows from experience.

He may be completely right in this analysis, particularly if his staff is perceptive and highly qualified—or he may misread the situation completely due to an honest misinterpretation of reality. But even if systematic evaluation tells the director exactly what he has concluded already, the effort reinforces his conclusions. Most careful evaluations, however, can tell even the most astute observer something of value that can improve the program.

A project team—politically speaking—may find some advantages in "going on faith" as compared to "going on evidence." Who can refute their faith and who has more experience than they? If they say or imply with conviction yet with openness that the project is developing very nicely, who can refute them? If the project team, and particularly the director, has the trust of his constituency, he may find "faith" is all that is required.

But he should know better! Education has too long bumped along on hope and faith, as important as these are, without serious effort to systematically ask and tackle the hard questions of evaluation. PACE projects should be expected to show the way in exemplinariness—and this includes evaluation.

Inadequate Evaluative Techniques

Egon Guba outlines three basic lacks: "(1) the lack of adequate theory, models, and designs to guide evaluative activity (as evidenced by the fact that even the 'expert' consultants do a bad job), (2) the lack of trained personnel (as evidenced by the fact that even the most rudimentary principles of the game are consistently violated), and (3) the lack of appropriate data collection techniques and data processing facilities (as evidenced by the heavy reliance on standardized tests or informal judgments)." (Guba, A-229 And Robert Stake writes: "New techniques of observation and judgment need to be developed. In fact, we need a new technology of educational evaluation. We need new paradigms, new methods, and new findings to help the buyer beware, to help the teacher capitalize on new devices, to help the developer create new materials, and to help all of us to understand the changing educational enterprise."[a]

Related to inadequate evaluative techniques is the broader problem of defining evaluation. Educators traditionally have defined evaluation as measurement of outcome (product) or judgmental with very little attention being given to what Egon Guba and Dan Stufflebeam have called for context, input, process, and product evaluation. The 1966 ASCD yearbook entitled Evaluation as Feedback and Guide reflects the newer concept of evaluation. With respect to PACE proposals, Robert Havighurst observes: "There is almost nothing to show how the project will be evaluated while it is in progress, so that it can be improved, errors corrected, and progress reports made. This is especially important in innovative work, where one must expect to learn from experience as the project progresses. A periodic stock-taking is desirable, using local staff and community committees as well as outside consultants." (Havighurst, A-77)

Inadequate techniques and procedures for evaluation seriously impede progress in PACE and elsewhere, yet shortcomings relating to the "people problem" are probably more serious.

---

[a] Robert E. Stake, "Toward a Technology for the Evaluation of Educational Programs" In Perspectives of Curriculum Evaluation. Chicago: Rand McNally and Company, 1967, p. 3.

Inadequate Teacher Preparation

Incredibly little has been done in developing evaluation specialists in our colleges of education, (The UCLA Research and Development Center, The Evaluation Center at Ohio State University, and the Center for Instructional Research and Curriculum Evaluation at the University of Illinois are notable exceptions.) Until the need for evaluation specialists becomes a problem of teacher education also, significant progress cannot be expected.

Several universities need to commence graduate concentrations for developing specialists in evaluation. An increasing number of these programs need to focus solely on evaluation, with general work in the processes of educational change. These programs should not attempt to develop better teachers or curriculum workers or administrators. Graduate programs for colleges of education have shunned specialization in evaluation and the change processes but these areas inevitably will be forced upon them by the nature of demand. (This position does not interprete specialization as a narrowness of focus to the exclusion of breadth. Quite to the contrary, the specialist of the future will have a better grasp of multidisciplinary aspects of education than the generalists today who usually does not step far beyond the offerings of his subject area.)

Shortage of Qualified Specialists

The demand for evaluators in the future appears destined to far outstrip the supply. In this respect, the outlook is dismal indeed. To quote from a study by Hopkins and Clark:

> Conservative manpower projection figures for the next five years are staggering. Disregarding all government support programs other than OE, all private foundations programs, and the inevitable stimulation of state and local activity in R and D which will result as an offshoot from the Federal support programs, that is, concentrating only on predicable growth of the OE's R and D support, education will need a hard-core R and D personnel pool of 130,000 by FY '72. In terms of the spread across R, D, and D, and translated into F.T.E. (full time employees), the demand picture will be roughly as follows:

	Percent	Full-time employees
Research	14	9,200
Development	46	29,700
Diffusion	40	[1] 25,500

> [1] D. L. Clark and J. E. Hopkins. "Preliminary Estimates of Research, Development and Diffusion Personnel Required in Education." Special project memorandum, Sept 1, 1966, pp. 15-16.*/

Two consultants—William Alexander and Elliot Eisner—allude to the shortage of qualified specialists in evaluation. Eisner observes that "given the dearth of competent people in the field of educational evaluation this lack of attention to the assessment of different populations and the relationship among data secured from these populations is not surprising." (Eisner, A-58 And Alexander notes that ". . . as guidelines require greater skill and precision in evaluation, the problems of local educational groups become more acute. Obviously a minority of school systems have enough personnel with adequate training and experience in evaluation to prepare proposals that include comprehensive and sophisticated evaluation plans." (Alexander, A-10)

---

*/ Quoted in Notes and Working Papers on Administration of Programs, Title III, ESEA, "Administrations of Programs," (U. S. Government Printing Office), April, 1967, p. 69-70.

With the evident shortage, what can be done about it? Several consultants have made recommendations for action. Those of Finn, Guba, and Schramm seem particularly relevant in considering positive steps.

Recommendation No. 1 is drawn from Guba's paper and the rationale for it can be found there:

I. THERE SHOULD BE ESTABLISHED:
   A. A NATIONAL LABORATORY FOR THE STUDY OF EVALUATION
   B. A NATIONAL INFORMATION CENTER FOR EVALUATION
   C. A NATIONAL GRADUATE SCHOOL FOR EDUCATIONAL EVALUATION (Guba, A-231)

Guba is saying, in essence, if the Federal government is really serious about evaluation, then it should facilitate establishment of institutions that will assist on a long term basis in the critical problems associated with improving evaluation.

John Letson in his chapter recommends the conduct of "small seminars to assist local agencies, which are operating programs with similar objectives, in the development and refinement of their evaluative plans." (Letson, A-128)

Related to both the Guba and Letson recommendations, James Finn recommends:

1. Title III funds be used to set up a series of regional evaluation centers throughout the United States designed to provide training and assistance to local education agencies.

2. The function of these centers be to provide advice, training and services and, particularly, to diffuse the general idea of the importance, usefulness and nature of a high-quality evaluation system.

3. It be understood that the evaluation centers are only persuasive and helpful in nature and that, if an educational agency chooses not to respond, it be allowed to without penalty—actual or implied.

4. These centers also engage in a certain amount of applied and field research with the purpose of developing viable and variable evaluation procedures which can embrace all types of evaluation needs and purposes.

5. A back-up national board be set up to assist the centers and the U.S.O.E. and Congress. This board would have the following functions:

    a. Locate and rotate manpower between the centers.

    b. Act as the assembling agency for results which ought to be diffused as the communication agency between the centers.

    c. Engage in broad scope research and development studies in. . . evaluation.

d. Provide an information source.

e. Relate to and diffuse information. . . (Finn,A-200)

Several other consultants also offered positive steps for improvement. Taken as a whole, the consultants are calling for innovation and creativity in meeting the shortage of evaluation specialists. These comments lead to recommendation No. 2, which is based upon one by Ira Singer:

II. EVALUATION COMPETENCE FOR ESEA TITLE III SHOULD BE DEVELOPED THROUGH SUMMER INSTITUES, INSERVICE EDUCATIONAL TELEVISION FEATURING VIDEOTAPES DEVELOPED FOR NATIONAL USAGE, AND SPECIAL MANUALS.

NDEA institutes might well be developed on PACE evaluation. Twelve, six week summer programs might make a significant dent in the PACE evaluation, especially if this cadre would be on call to help other PACE projects.

CHAPTER III

HOW EFFECTIVE IS PACE EVALUATION?

The 19 consultants made a total of 114 comments about project evaluation: 78 were negative, six were positive, and 30 were general. It is interesting to note that evaluation also was the issue mentioned most frequently in the first national evaluation of PACE.

Without exception, the special consultants criticized the effectiveness of PACE evaluation. Their negative comments ranged from very strong to mild, but even the mildest critique was decidely negative. Typical of most consultants' reactions were those by Wilbur Schramm, Harry Passow, and Elliot Eisner. Schramm wrote: "One comes away from reading these evaluation sections rather certain that (a) evaluation means very little in these project plans; (b) little thought has been given to what constitutes success—at least behaviorally—in these projects; and (c) few of them have had the benefit of much attention from a research man." (Schramm,A-150) Passow concluded: "The impression one gains from reading the proposals is that, with few exceptions, the evaluation section is written hopefully to satisfy the Office of Education's requirements—not because evaluation is perceived as having any intrinsic value for the program itself." (Passow, A-134) And Eisner found "the evaluation sections are weak, nonanalytic and frequently use inappropriate instruments to assess poorly specified outcomes." (Eisner, A-60)

All projects are not evaluating poorly. Arthur Hitchcock found that "one project has reached the point in evaluation procedures of a feedback, change, and decision-making process. Two projects had a model for their education," (Hitchcock,A-89) And William Alexander said that "three proposals included somewhat comprehensive and unique patterns of evaluation." (Alexander, A-4)

Very likely more and better evaluation is taking place on project sites than evidenced their project proposals. Most project directors and their staffs started "cold" on evaluation but a decided warming trend has developed. Conferences such as the three Hawaiian Seminars on Innovation and many others have induced some knowledge about evaluation. Also, almost any highly successful administrator is constantly evaluating—and in the context of feedback-guide-judgment. But the exception does not make the rule; flying by "the seat-of-the pants" is not for jets any more than simple wisdom is appropriate for evaluating complex educational problems.

Comparison of 1966 and 1967 Proposals

The 19 special consultants reviewed evaluation sections, or at least searched for such sections, in a random sampling of 379 proposals. Of this number, 101 were submitted in 1966, 190 in 1967, and 88 were not identified by year by the special consultants. The 1967 proposals were written from the mid-1966 guidelines which required a section on evaluation.

What affect did this requirement have?

Thirteen comments by special consultants pertained to the comparison between 1966 and 1967 and in all cases some or much improvement was noticed. John Letson wrote: "The increased emphasis given to the evaluative process, as evidenced in the successive editions of PACE, is reflected in the designs of the evaluative sections of the sampled proposals. More proposals approved in 1967 received an above average rating than those approved prior to 1967." (Letson, A-122) Arthur Hitchcock wrote "Earlier ones tended toward generalities, descriptive data, and a small variety of instruments, frequently limited to achievement tests. The 1967 ones tend toward:

1. Displaying the language of evaluation.
2. Evidencing an acceptance of the evaluation concept of measuring progress toward objectives.
3. Using a greater variety of instruments within the evaluation of a single project."

(Hitchcock, A-89)

And Lloyd Dun found: "With the advent of the Manual for Project Applications and Grantees. . . and with a Chapter V outlining the evaluation requirements, the attention paid to this topic changed dramatically. While one can question the vigorousness of both the stimulus and the response, and it is clearly evident that setting expectancies on the part of the grant-awarding agency did elicit attention and concern for this matter on the part of the grant-seeking systems." (Dunn, A-41)

It is clear that requiring proposals to consider inclusion of provisions for evaluation has brought about some improvement. The point at which relationship between structure becomes an impediment to innovation and creativity is an interesting one, however. James Finn raises this concern when he mentions "that this movement toward system in evaluation also may not lead to improvement. For, in order for an evaluation system to be applied across the country, it is necessary first to institutionalize it; this is to say that, unless other means are invented, the evaluation system must be initiated, monitored and controlled by a bureaucratic system. Institutionalization of the evaluation process could destroy the innovative possibilities of Title III." (Finn, A-173)

One should consider the evaluative weaknesses of PACE in the more general context of American education. Evaluation is also very weak in ESEA Title I proposals, and philanthopic foundations have done little in this area. Blame for evaluation deficiencies of PACE rest with larger deficiencies, and shortcomings found in the PACE proposals and projects reflect the larger dimension.

CHAPTER IV

MAJOR PROBLEMS IN PROPOSALS

This chapter focuses upon those evaluation problems identified most frequently by the 19 special consultants. Three major problems will be discussed, and some solutions to meet the problems will be suggested.

Three Major Problems

1. Objectives

The special consultants made 33 comments about the objectives. In general, their comments might be clustered around two criticisms. The first general criticism related to lack of general understanding of behavioral objectives. Don Davies writes that "the fact that behavioral objectives are 'in' these days is reflected in many of the proposals. However, it is clear that the educators writing these proposals have incomplete, inadequate, or distorted understanding of what behavioral objectives are and how they are related to measurement and evaluation." (Davies, A-30) Maurie Hillson finds that "some difficulties and problems of evaluation now may be due to the fact that the statements made concerning aims and objectives are somewhat grandiose." (Hillson, A-85)

A second general criticism concerned the relationship between the stated objectives and the selected evaluative procedures. Don Bushnell notes that "all projects proposing use of computer assisted instruction (CAI) in mathematics and science do not evaluate results in terms of all objectives they proposed." (Bushnell, A-22) On this point, Dorothy Fraser writes: ". . . evaluative measures are related to project objectives, but in nine of 12 proposals there is little or no effort to relate a particular project objective to particular evaluation procedures." (Fraser,. A-67) and Glenn Blough noted that "the plans for evaluation are not 'geared into' the objectives indicated. . ." (Blough, A-16)

Robert Stake states that "to understand the Title III operation and to ascertain its value, we are obligated to identify goals, ascertain priorities, reveal the dynamics of changing priorities, and provide information for decisions about new goals and priorities." (Stake,A-209)

Again, the proposals reflect a serious weakness in American education; namely, giving lip service to objectives. The tendency is to develop an idea in terms of bringing about some improvements, but rarely do project developers force themselves in the difficult position of making precise decisions about objectives. But this initial step is essential in order for effective evaluation. This position leads to Recommendation No. 3.

III. EVERY PROPOSAL SHOULD AMPLY DEMONSTRATE THAT OBJECTIVES HAVE BEEN CONSIDERED AT THE GENERAL AND SPECIFIC LEVELS

Furthermore, learning or behavioral objectives should be related to program activities, and the types of evaluation used should be related to activities. Different activities require different types of evaluation—a tantalogy well illustrated in a very useful booklet entitled A Guide to Assessment and Evaluation Procedures[a].

At this point, however, a caution flag is raised with respect to uncritical acceptance of objectives. The stated need to begin with "behavioral objectives" has almost reached slogan proportions in American education.

Few will deny that greater attention needs to be given to behavioral and content objectives, but several factors should temper action. Elliot Eisner has spelled out four limitations of using educational objectives in curriculum theory, and these seem quite apropos to some PACE activities. He writes: "First it (curriculum theory) has not sufficiently emphasized the extent to which the prediction of educational outcomes cannot be made with accuracy. Second, it has not discussed the ways in which the subject matter affects precision in stating educational objectives. For instance, it is much easier in mathematics and spelling than in arts and social studies.) Third it has confused the use of educational objectives as a standard for meaurement when in some areas it can be used only as a criterion for judgment. (Again, measurement is possible in mathematics and spelling, for example, while judgment applies in the arts and to some extent in the social studies.) Four, it has not distinguished between the logical requirements of relating means to ends in the curriculum as a product and the psychological conditions useful for constructing curricula.[a/] (In other words, as James MacDonald has pointed out, "the teacher asks a fundamentally different question from 'What am I trying to accomplish?' The teacher asks, "What am I going to do?' and out of doing comes accomplishment.")[b/]

---

[a] The New England Educational Asssessment Project. A Guide to Assessment and Evaluation Procedures. Providence, Rhode Island: the Project, 1967, 28 p.

2. Procedures for Evaluation

The 19 consultants made 87 comments about procedures for evaluation. Figure No. 1 lists 37 different types of evaluative procedures that were listed in 379 projects, and Figure No. 2 draws upon data gathered 18 months earlier.

FIGURE NO. 1

TYPES OF EVALUATIVE PROCEDURES

Types of Evaluations Listed in a Sampling of 413 PACE Projects	Frequency of Mention	Percentage of Total Number of Types of Evaluation Mentioned
Tests		
Standardized (commercial)*	78	18.9
Locally constructed	5	1.2
Consultant assistance	55	13.3
Research design		
Pre-post test	34	8.2
Experiemental (and control) groups	10	2.4
Surveys and questionnaires	38	9.2
Informal judgments	25	6.0

[a/] Elliott W. Eisner, "Educational Objectives: Help or Hinderance," An expanded version of a paper presented at the 50th Annual meeting of the American Education Research Association, Chicago, 1966, p. 10.

[b/] James MacDonald, "Myths About Instruction," Educational Leadership 22:613-14; May 1965.

FIGURE NO. 1—Continued

Types of Evaluations Listed in a Sampling of 413 PACE Projects	Frequency of Mention	Percentage of Total Number of Types of Evaluation Mentioned
Interviews	23	5.6
Increased participation and use of facilities	17	4.1
Case studies	16	3.4
Classroom observations	12	2.9
Attitude scales	11	2.7
Follow-up studies	11	2.7
Student achievement (success)	8	1.9
Comparative studies	7	1.7
Ratings	6	1.5
Teacher performance	6	1.5
School administrators' reactions	7	1.7
Parent reactions	5	1.2
Directors' reports	5	1.2
Checklist	4	.97
Teacher or administrator self evaluation	4	.97
Audio or visual record	3	.72
Conferences	3	.72
Advisory committees, interaction analysis, cost studies, visitors reactions, teachers diaries	2	.49
Student self-evaluation, student interest, pupil diaries, teacher achievement test, staff meetings, teacher turnover, pilot program, stimulation techniques, sociograms, interest inventories	1	.24

---

[*] Those listed were: California Achievement Test, Flanders Interaction Analysis, Holpines Climate Index, Iowa Test of Basic Skills, Iowa Test of Educational Development, Metropolitan Achievement Test, Stanford Achievement Test, Survey of Reading Practices in Georgia Schools, Teaching-Learning Process Analysis Inventory, and Watson-Glazer Test of Critical Thinking.

FIGURE NO. 2

EVALUATION METHODS PROPOSED IN 138 OF THE 174 OPERATIONAL PROJECTS APPROVED IN THE 1ST AND 2ND ROUNDS

		Number	Percent
1.	Standardized tests	92	66.7
2.	Teacher (librarian) evaluation	67	48.5
3.	Outside special evaluators	37	26.6
4.	Conferences, interviews, questionnaires	34	24.6
5.	Directors evaluation	34	24.8
6.	Subjective tests	29	21.0
7.	Increasing achievement (basic skills)	[1] 22	15.9
8.	Student reaction	26	18.8
9.	Parent attitudes and community responses	21	15.2
10.	Increase, voluntary participation	20	14.5
11.	By consultants	19	13.7
12.	Annual and other reports	16	13.7
13.	Number using library or materials	15	10.9
14.	State departments of education	15	10.9
15.	Student attitudes, improve self and society	14	10.2
16.	Attendance, increase	12	8.69
17.	Check lists, inventory, surveys, polls	9	6.5
18.	Fewer dropouts	8	5.7
19.	Computer data	7	5.07
20.	By movies and taped records	7	5.07

[1] Increasing achievement (presumably the results of objective tests).

Certainly there is no lack of variety in the types of evaluative procedures used, but there is serious question about how these are used. Joseph Rubin cautions that "evaluation must not be confused with testing. The two must be clearly distinguished." (Rubin, A-145)

Three weaknesses might be said to summarize the observations of the special consultants.

1. Best procedures are not chosen. A number of consultants made this point. For example, Dorothy Fraser writes: "There is little indication that the current literature dealing with evaluation in the social studies has been examined as a basis for planning for project assessment (Fraser, A-69); Elliot Eisner notes: "Another characteristic. . . is a tendency to employ published but inappropriate instruments for purposes of evaluation." (Eisner, A-59); and Egon Guba finds that "evaluators neglect many useful techniques that are available." (Guba, A-228)

The use of experimental designs for evaluation of PACE projects is appropriate in some instances, but a tendency to "contract out" development of an experimental design is questionable in many others. Two weaknesses of the experimental design approach are pointed out by Daniel Stufflebeam, special advisor to the National PACE Evaluation Team: (1) "the application of experimental design. . . (to evaluation problems) conflicts with the principle that evaluation should facilitate the continued improvement of a program. Under both design and analysis principles of experimentation, treatments cannot be altered in process if the data about differences between treatments are to be unequivocal. Thus, the treatment must accommodate the design rather than promote changes in the treatment. . . A second flaw in the experimental design type of evaluation is that it is useful as a judgmental device but almost useless as a decision-making device."[a]

2. Selected procedures are not closely related to activities. Glenn Blough writes that "a common characteristic of nearly every statement concerning the plan for evaluation is vagueness. Perhaps 'Let's indicate that we will evaluate and then solve the problem of how we will do so when the time comes,' summarizes the situation." (Blough, A-13)

Evaluation procedures listed in proposals rarely are related to specific activities. For example, what is to be measured; the acquisition of specified facts, positive changes in specified attitudes, or improvement in specified performances? For acquisition of facts, one might turn to established objective tests or to locally designed ones; for changes in attitudes one might resort to a questionnaire, structured interview, or case study; and for improvement in performance one might select or develop a rating scale, checklist, or standardized tests. These examples simply illustrate the necessity of fitting the evaluative procedure to the activity.

3. Expert advice in developing procedures is not used. Arthur Hitchcock finds that "the better evaluations had outside consultative help," (Hitchcock, A-95) but evaluation consultants were in short evidence in the proposals. Wilbur Schramm has "the uneasy feeling

---

[a] Daniel L. Stufflebeam. "The Use and Abuse of Evaluation in Title III," An Address Delivered at the National Seminar on Innovation, Honolulu, Hawaii, July, 1967, p. 7.

that a research man hasn't been involved in these proposals, except perhaps perfunctorily, and isn't likely to be. . ." (Schramm, A-151)

While many proposals plan to use consultants, these are rarely evaluation or research specialists. One also has the feeling that some project directors believe any good educator can also evaluate. This naive approach to evaluation is appalling, and it is responsible in large measure for the low esteem of evaluation and research.

A mania for "hard" data is a general problem that should be mentioned. "Hard" data refers to that evidence gathered from standardized test results and/or from experimental research designs. A few Chi squares, multiple regressions, and computer analyses are badges of respectability and somehow are supposed to result in better product evaluation.

But how hard is hard data? In the final analysis, all standardized tests reflect the biases of the inventer and all research designs must decide upon what to measure. In other words, objective evidence is based upon initial subjective judgments with respect to priorities and values.

Of course standardized test scores and research designs should be used wherever they will satisfy established purposes, but surveys, interviews, case studies, observations and the like also can be entirely valid and important sources of evaluative data, depending upon how they are used. Recommendation No. 4 relates to this point.

IV. PROCEDURES FOR EVALUATION SHOULD CLOSELY REFLECT THE NATURE OF THE TASK OR PROJECT TO BE EVALUATED

The current interest in cost effectiveness and cost benefit studies has prompted greater attention to hard data, and this attention on the whole is desirable, but PACE directors must not try to force hard data procedures upon unlikely situations. Robert Havighurst, in commenting on the problems of evaluating Supplementary Educational Centers, points out that "the programs of the Centers tends to be broad, and rather vaguely defined. They usually propose to create new courses of instruction with new teaching materials, or to train teachers and counselors for new roles. They do not lend themselves to an experimental design, with experiemental and control groups of students and statistical tests of various hypotheses." (Havighurst, A-32)

3. Amount Budgeted for Evaluation

Thirteen comments by the special consultants related to the amount budgeted for evaluation. William Alexander estimated that the proposals that he examined allocated approximately two percent of their total budgets for evaluation, but one is forced to extrapolate this figure from several places in the budget. (Alexander, A-1) Other consultants found little evidence of serious budgetary commitment to evaluation. Don Davies, for example, finds that "evaluation does not appear as a separate budget item in any proposal." (Davies, A-35) Dorothy Fraser did find that "proposals that did include budget items for evaluation tended to be more definite in their plans for evaluating the project than those that did not." (Fraser, A-67)

Recommendation No. 5 relates to the problem of amount budgeted.

V. EVERY PACE PROPOSAL SHOULD BE REQUIRED TO HAVE A SEPARATE BUDGET ITEM FOR EVALUATION, AND THE AMOUNT OF THIS FIGURE SHOULD NOT BE LESS THAN FIVE PERCENT OF THE TOTAL BUDGET.

Very little or no budgetary commitment to evaluation results in very little or nothing. The return expected is directly related to the investment made. Proposals simply must have a well defined and adequate evaluative expenditure to expect sound results. The five percent figure is not based upon research but upon experience with a few proposals that seem to have an adequate evaluation scheme. Some evaluation schemes go up to ten percent of the total budget.

Solutions to Meet the Problems

Growing out of the discussion thusfar, this section will include several recommendations for improving PACE evaluation. Recommendation No. 6 is a general one related to revision of guidelines.

VI. NEW GUIDELINES NEED TO BE DEVELOPED BY THE USOE THAT WILL STRENGTHEN PROJECT ASSESSMENT.

The special consultants almost uniformly call for this strengthening, but not at the expense of squeezing out creativity or innovativeness. James Finn expresses this concern when he states: ". . . while the analyses of the experts. . . of the evaluation process are impressive and potentially fruitful, is it possible that they have, in fact, over-analyzed the process and, in doing so, slipped into the same trap that the conventional educational research man does when he attempts to apply controlled research techniques to evaluation processes operating under field conditions? . . . Have, in fact, these analyses departed from operational reality, at least in the sense that the practitioner would not know what to do with them? And, if one or more of these models was frozen into enforced guidelines, would this not result only in bureaucratic paper?" (Finn, A-181)

The problem of guidelines is complicated by the widely varying differences among PACE projects, ranging from a single task, single school project with modest financing to multi-task, multi-district effort with several hundred thousand dollar expenditure. The objectives established for these two projects would be vastly different and therefore so should be the evaluative procedures.

The problem of guidelines is further complicated by differences in types of activities. Elliot Eisner, for example, points out that "it should be made clear at the outset that the evaluation of learning in the cultural arts is a task beset with a variety of special difficulties. . .

The context in which evaluations in the cultural arts reside is one that tends to have little disposition toward objective evaluation and few instruments appropriate for evaluation when unique objectives are formulated." (Eisner, A-56)

The application of Recommendation No. 6 needs to skirt between the Charybdis of specificity and the Scylla of standardization without losing the creativity and innovativeness that characterizes PACE.

Recommendation No. 7 relates to a further step toward developing materials to assist local project directors as well as state coordinators.

VII. THE USOE SHOULD DEVELOP SIMPLE YET ACADEMICALLY SOUND MATERIALS ON THE THEORY AND PRACTICE OF EVALUATION THAT WILL PROVIDE CONCERETE ASSISTANCE TO PROJECT DIRECTORS

Financing the ESEA Title III largely will be turned over to the States within two years yet the Commissioner will maintain the responsibility for approving all State plans, and within these could be required detailed evidence that evaluation was being given serious attention within the State. The availability of sound materials would have an important role in improvement of projects within respective states.

Recommendation No. 8 is related to Recommendations No. 1 and 2 but No. 8 zeroes more specifically on a suggestion where the earlier ones tended to be more general.

VIII. SPECIAL ONE OR TWO WEEK SEMINARS ON THEORY AND PRACTICE OF EVALUATION SHOULD BE DEVELOPED SPECIFICALLY TO PROVIDE ASSISTANCE TO PACE PERSONNEL

Several industrial concerns have invested sizable human and material resources into developing intensive short courses for their personnel. These courses, often one week in length, are characterized by very careful planning; small groups, usually less than 40; rather specific objectives; and evaluation and follow-up procedures.

A Call for Perspective

Finally, it is important that one keeps perspective on evaluation. We must not conclude that because a project has not developed an acceptable scheme for evaluation that it necessarily will be a poor project. The application of this standard would mean that PACE was a "bust". Of course this is not so. In spite of the weaknesses and problems, PACE remains a "swinging" title. Much of this success rests with the project directors in the field, and these individuals and their staffs do represent a positive and dynamic force in American education.

PACE projects as a whole are doing well in spite of quite inadequate evaluation procedures. Many projects, however, can improve only modestly without serious attention to evaluation. This area offers great potential for significant improvement on part of all ongoing programs.

CHAPTER V

REPORTS BY SPECIAL CONSULTANTS

Each of the 19 special consultants was asked to study approximately ten approved PACE proposals that more or less focused upon his area of speciality. (A total of 379 were examined.) Each consultant was asked to judge the adequacy of evaluation procedures for these ten projects. The effort was not to evaluate the proposals as a whole but only the evaluation sections.

The following reports represent the independent scholarship of the consultants. Each report has analyzed the effectiveness of evaluation procedures, and almost all reports include recommendations.

CURRICULUM DEVELOPMENT[*/]

The 34 Title III proposals reviewed herein related generally to curriculum and instructional improvement. Included were proposals for:

> Planning and operating various types of supplementary service centers to improve many aspects of curriculum and instruction
>
> Operating curriculum planning centers for individual schools (especially frequent, middle schools) and school districts
>
> Operating various individualized instructional services and independent study plans
>
> A miscellany including a project in American history instruction, an occupational counseling center, the use of a male teacher in primary grades, and a comprehensive in-service training program

As to intent, 14 proposals were for planning grants with $1,465,899.10 requested; 16 for operational grants, $3,486,803.38 requested; and 4 combining planning and operational grants, $1,839,246.80 requested. Thus the 34 grants anticipated a total of $6,791,949.28 in federal funds over the years for which budgets were proposed. Of this total amount, identifiable budgetary items for evaluation in planning amounted to $5,700.00, and for evaluation in operation, $121,765.00—for a total of specific items designated for evaluation of $127,465.00, or approximately two percent of the total funds budgeted. However, the lack of specificity in many sections on evaluation, and in some budget

---

[*/] Prepared by William M Alexander, Professor of Education and Director, Institute for Curriculum Improvement, University of Florida.

provisions, plus the uncertainty of continuation grants preclude a judgment based on such figures alone, relatively small as the evaluation budget total seems.

Description of Procedures and Approaches to Evaluation

Of the 34 proposals received, 12 were considered by the reviewer as lacking identifiable provisions for evaluation. Of these, nine were for planning grants, and the other three had been submitted before the January, 1967, submission date when evaluation sections were generally expected in operational grant proposals. The remainder of this review is therefore devoted to evaluation provisions in the 22 proposals having such provisions.

Six of these 22 proposals were for planning grants. Procedures of evaluation proposed in these included the following:

> Longitudinal evaluation of teacher behavior, pupil behavior, and pupil achievement (instruments not indicated)
>
> "Trial run" of a planned program of instruction, during a special summer program
>
> Use of a group including "professional evaluators" and research specialists to evaluate plans
>
> Employment of an outside evaluator to assess curriculum planning activities
>
> Employment of a research assistant to design studies, carry on field tests, and analyze data
>
> Organization of a pilot program to be evaluated under the direction of an evaluation committee

A pre- and post-testing program in conjunction with developing a different program of instruction in one subject

With the exception of the last category, procedural descriptions are quite general in nature. The proposals say, in effect, that the results of planning will be evaluated by some group or individual without specifying very clearly the nature of the process. By contrast, one instructional program in one subject (American history) quite specifically details development and use of instruments for pre- and post-testing of the program developed, and other product evaluation procedures.

The 16 operational grant proposals included a variety of evaluation procedures. Among them the following were noted as individual approaches of a somewhat usual nature, although in some cases proposed for use in an innovative way:

Follow-up studies of various groups

Pre- and post-testing of children, teachers, and other groups, employing a variety of instruments

Judgments of participants, consultants, staff members, and others

Individual research studies of various innovative projects in model school programs

Use of an outside organization to make a comprehensive evaluation (usually described in very general terms)

Records of teacher participation in inservice and other projects, and of changes reported as affected thereby

Comparison of high school course offerings before and at end of a stimulation-type project

Comparison of pupil achievement in a pilot and control situation

Reactions of visitors to a program

Records of student and community participation in a cultural activities program

Interviews and questionnaires for inventories of attitudes (sometimes used before and after)

Evaluators' review of tapes and other recordings of activities

Records of use of instructional materials and methods

Three proposals included somewhat comprehensive and unique patterns of evaluation. Excerpts from these proposals follow to illustrate these patterns:

1. From a project on independent study in a high school:

   (The project) proposes to extend and test the impact of increased independent study commitments on the appropriateness, feasibility, and success of the plans and decisions of members of these self-designated student groups. . . . .

   In addition to the evaluative criteria used in previous studies (grades, tests, and post high-school success), (the project) will monitor the learning attitudes and learning behaviors of members of these self-designated student groups as they are involved with an increasing variety of independent study media and opportunities. Systematic observations, structured interviews, self-reports and study-log samplings will be used to supplement

> the "hard data" provided by graded student success in classrooms and by performance on standardized and locally-constructed tests within basic subject areas. Data derived from the assessment of members of these self-designated groups within (the project school) will be compared with comparable data derived from similarly self-designated groups in the other two high schools of the district who will not have had extended opportunities for independent study, but who will have full benefit of common curricular and guidance systems operational within the district.

2. A proposal for a regional educational services center provided for the establishment of an evaluation and research component and described specific procedures for evaluating each of three types of the center's services. For example, one element, "selective dissemination of information," of the center's information service component would be evaluated through these means:

> Quantitative evaluation will be prominent. Records of numbers of requests will be maintained as well as categories of requests. The degree of sophistication of inquiries will be determined from these records. Daily logs, using data coding to be developed, will assist in the above task. Checks with consultants will be made to ascertain their utilization of the service. Follow-up samplings will be made to determine the degree of change effected by usage of the service. Items of concern will be: who utilized the service, were they benefited, did the information affect students, did it affect teacher behavior. The information derived from this technique will be used to institute necessary revisions of this service.

3. A cooperative evaluation venture between a school system

and a university evaluation center is partially described in this excerpt:

> The proposed strategy is to collaborate with the________________University Evaluation Center in the design and implementation of evaluations for our Title III project. The Center has assisted in preparing this evaluation design and will direct its implementation. Our school, however, will provide staff members to serve on a school-and-university evaluation team. They will be trained in principles of evaluation by the Evaluation Center and will subsequently collect the basic data for evaluations. The Evaluation Center will provide instrument construction and data analysis services.
>
> . . . . . The Evaluation Center will, in effect, be conducting "on the job" training for us. We will also be receiving the benefit of the Center's experience and facilities in conducting the evaluations that are required now. Further, such collaboration will afford the Evaluation Center an opportunity to study the evaluation process in a "real-world" context. This should facilitate the production of generalized evaluation designs which meet the requirements of school enterprises.

Analysis

The classification of types of evaluation used by Guba[1/] and Stufflebeam is adopted for this analysis, with an additional statement on budgetary provisions for evaluation in the proposals. Only the 16 operational proposals selected as above are considered in this section.

[1/] Egon G. Guba, "Evaluation and the Process of Change," in Richard I. Miller, ed., Catalyst for Change: A National Study of ESEA Title III (PACE) (Washington, D. C.: Government Printing Office, April, 1967), pp. 307-308.

Context evaluation: Several of these projects had been preceded by planning grants in which data had been compiled that were used to justify or explain the need for the operational grant. The proposal guidelines require statements as to needs and planning, and in some proposals these sections present sufficient data to describe the background of the project. In some cases the characteristics of the population to be served are stated with enough factual detail to define clearly this aspect of the context. In other cases the results of prior planning and needs identification are lacking; in most of the proposals the section on planning, for example, merely lists the persons and groups involved with little attention to what they have done (or will do).

Input evaluation: Description of means that had been employed in planning or would be in operation to assess in advance the proposed activity, is generally lacking. In a very few proposals the description of prior planning phases does include some report of the consideration of alternative solutions of the central problem. Only one proposal indicated that related research studies would be searched out and utilized. Mention of pilot study approaches was noted in only three proposals.

This reviewer believes that a statement of the project's rationale should defend a projected activity as more promising and feasible than other possible approaches to the problem. On this criterion, not more than three of the 16 proposals seem satisfactory.

Process evaluation: Most proposals project planning and evaluation activities which could provide ready sources of feedback data that could be used to modify the project in operation. Only one proposal included a specific section on "process evaluation," although two others also described in some detail how feedback would be compiled and used in the course of the project.

Several projects were to include substantial efforts to effect curriculum and instructional change through inservice education activities for the staff, but these proposals generally failed to describe how related data, including teacher reactions, would be collected and used in the course of the project to modify the inservice program. The proposals also generally failed to specify how, if at all, the deliberations of planning and evaluation bodies would be used to modify project activities.

Product evaluation: The evaluation sections of the proposals are focused on product evaluation. The variety of procedures were described earlier in this paper, and certain approaches considered exemplary by this reviewer were excerpted.

In general, the clarity and relevance of plans for evaluating project outcomes seem much greater in these 16 proposals than in earlier ones read by this reviewer. Only one of the 16 had such a general statement as to indicate the absence of any evaluation plan. It was disappointing to find relatively little suggestion as to means of evaluating the nature and extent of curriculum change in the individual schools served by an educational services center. Changes in curriculum content, curriculum offerings, provisions for curriculum differentiation, and other aspects of a school program, seem as susceptible to analysis and evaluation as inservice education opportunities, use of instructional resources, and other frequent evalution foci.

Budgetary provisions: Although each of these 16 proposals included evaluation sections, most of them having relatively detailed plans for product evaluation, five proposals included no budget item that could be identified as an evaluation cost. Of the other 11, the identifiable budget items were of very small amounts in some projects: $802 and $600, for example, in two projects each requesting grants of six-digit size. The most easily identifiable budget items were for projects allocating substantial amounts for evaluation services to be contracted ($24,000 in one proposal) or for evaluation personnel to be employed. However, proposals were also reviewed that anticipated using evaluation specialists but failed to designate budget items for these services. Perhaps this situation is due to confusion or change in guideline requirements.

Interpretation

This analysis indicates that evaluation procedures are more frequently included and in general of higher quality in the sample of the 1967 Title III proposals as compared with those of a year earlier. The need for further improvement remains obvious, however.

One problem seems to lie in the guidelines themselves. This problem was well exposed with relevant remedial recommendations in Guba's review of last year.[2] Hopefully, subsequent guidelines will be more explicit in their directions as to the evaluation proposals desired.

However, as guidelines require greater skill and precision in evaluation, the problems of local educational groups become more acute. Obviously a minority of school systems have enough personnel with adequate training and experience in evaluation to prepare proposals that include comprehensive and sophisticated evaluation plans.

Recommendations

In view of this analysis of 34 proposals related to curriculum development, the following recommendations are offered

1. IMPLEMENT INTO THE TITLE III PROGRAM AND ITS GUIDELINES THE RECOMMENDATIONS MADE BY GUBA (and of other members of the first-year study team as were relevant to this point), especially these two:

   (1) Provide adequate guidelines for the local proposer on the matter of evaluation.

---

[2] *Ibid.*, pp. 312-314.

(2) Help the local proposer understand the meaning and utility of each of the four kinds of evaluation: context, input, process, and product.[3]

In addition, guidelines as to budget preparation should give specific guidance as to how to designate budgetary items for evaluation.

2. LIBERALIZE BUDGET SPECIFICATIONS AND RESTRICTIONS TO ENCOURAGE SCHOOL DISTRICT-UNIVERSITY COOPERATION IN DEVELOPING PROGRAMS FOR EVALUATION THAT WILL SERVE THE DUAL PURPOSES OF PROJECT EVALUATION AND TRAINING IN EVALUATION.

As illustrated by one of the project excerpts cited earlier, the more exemplary evaluation proposals have involved or anticipated involvement of university evaluation centers. Properly and cooperatively developed, these arrangements could substantially expand the supply of trained evaluators as well as to increase the quality of project evaluations.

3. EMBODY IN TITLE III MANUALS AND GUIDELINES A THEORY AND PRACTICE OF EDUCATIONAL CHANGE WHICH PRESENTS EVALUATION AS AN INTEGRAL AND CONTINUING PHASE OF THE PROCESS.

---

[3] *Ibid*., p. 313.

To this end, the reviewer regards the following brief description[4/] as applicable:

I - Identify the need(s) for curriculum and instructional improvement

D - Determine possible innovations (change factors) which seem likely to satisfy the need(s)

E - Evaluate a chosen innovation in practice

A - Activate (by rejection, renovation and retrial, demonstration and/or diffusion) the results of evaluation

S - Stimulate continuously the foregoing steps in an ongoing change process

---

[4/] See William M. Alexander, "The Acceleration of Curriculum Change," in Richard I. Miller, ed., Perspectives on Educational Change (New York: Appleton-Century-Crofts, 1967), for a fuller development of this pattern.

SCIENCE EDUCATION[*/]

Twenty projects out of the 40 available constitute the material for this report. Ten were approved before the date of mid-1966 (until this date evaluation had not been emphasized); ten after.

The projects selected all relate to science teaching, in one form or other, in grades 1 through 12. Six projects are concerned with various phases of conservation, outdoor education and resource use, five with astronomy and the use of planetaria, three with marine study, three with mathematics and science seminars, one with an experience curriculum, one with the teaching of earth and space science and one with summer science experiences.

Analysis

Of the 10 projects approved after January 1967, only one failed to indicate some provision for evaluation. Of the 10 projects before, only two described any evaluation plans. Conclusions can not be drawn from this small sample but there is reason to believe that project planners did indeed take into account the directive to include some definite plans for evaluating their efforts.

A common characteristic of nearly every plan for evaluation, however, is vagueness. Perhaps, "Let's indicate that we will evaluate and then solve the problem of how we will do so when the time comes," summarizes the situation.

[*/] Prepared by Glenn O. Blough, Professor of Education, University of Maryland.

In drawing conclusions and describing the evaluative intentions of the projects examined, it seems important to analyze first the objectives as they are stated or implied in the project descriptions. Since effective evaluation must be based on objectives, this approach seems reasonable. Some projects listed the objectives, others implied them. Considering the 20 projects as a whole, more than one-half did not give specific objectives.

Listed in order of emphasis, eight projects gave teacher training as the chief objective, seven gave improving the teaching of subject matter, six described the preparation of curriculum materials (courses of study, bulletins, etc.) as the major objective, and an equal number indicated that developing interest in subject matter for pupils was their major concern. The following objectives were described as important in a few projects: develop teaching materials, provide for gifted pupils, provide first-hand experiences, supplement the present program, individualize instruction, and obtain more equipment.

The following methods of evaluation were suggested. These cannot be equated directly with objectives listed in the previous paragraph because in some cases the method of evaluation infers what is to be evaluated; in others it does not. Twelve projects are involved in the following discussion, or the number that indicated some methods of evaluation.

The most common method of evaluation is test evaluation. About one-half of the 12 indicated that pre-tests and post-tests would be given. Some indicated that standardized tests would be used, admitting that they did not know exactly where these tests would be obtained and whether or not such tests would actually measure the specific subject matter they planned to use. Only three projects indicated that they planned to use control groups in the testing program.

All 12 projects indicated that the evaluation (testing) would be done by teachers; three projects indicated that tests would be given at various critical points.

Eight projects indicated that questionnaires would be used. Samples, however, were not enclosed nor was there clear indication to whom the questionnaires would be sent.

Ten projects indicated that evaluation would include the number of pupils (or in some cases teachers or others) using the facilities, laboratories, centers, equipment, books, and the like. There was no indication that differentiation would be made between individuals who merely "signed the book" as having appeared and those that actually made extensive use of the facility.

Records and reports ranked next as a method of evaluation. Some indicated that these would be made periodically, others annually. This was followed by observation by consultants both from the school system and from outside it. Some included community leaders as observers, and one expected to include student opinion.

Among other methods of evaluation the following were mentioned once: surveys, follow-up study, visit by project director, and use of a general advisory committee. No details were included.

Summary

From the foregoing brief analysis, the following points seem evident: (1) Statements of evaluation plans increased after the mid-1966 date. (2) A variety of procedures are indicated. (3) None of the plans indicated seem innovative. (4) The description of evaluation plans does not make up a significant part of the project plan. (5) The "how" or the specific plan for carrying out the briefly stated possibilities are not included (this may be due to space limitation but I am inclined to believe that it stems from the fact that the plans have not been thought through very carefully). (6) The plans for evaluation are not "geared into" the objectives indicated. For example, inservice education ranks high in objectives and no mention is made of possible ways to evaluate it; development of interest seemed to be an important objective and there are no plans for even attempting to measure it. Granted the attainment of some of the objectives is difficult to assess, no progress can be made if there are no plans!

Recommendations

1. IT WOULD SEEM APPROPRIATE TO EXPECT GREATER EMPHASIS AND MORE DETAILS OF METHODS OF

PROCEDURES IN EVALUATION IN THE PROJECT PLANS WHEN THEY ARE SUBMITTED.

2. PROJECT EVALUATORS MIGHT BE URGED TO PLACE GREATER WEIGHT ON THE EVALUATION ASPECT OF THE PLANS AS THEY ARE SUBMITTED.

COMPUTER-MEDIATED PROGRAMS[*/]

The PACE proposals reviewed herein focus upon innovative programs which utilize computer technology as a means for realizing diverse educational goals. The goals of the projects reviewed are associated with the following productive programs:

1. The augmentation of mathematics and science curriculums through access to computer facilities
2. The automation of media centers
3. The establishment of a central data file to facilitate educational decision-making
4. The implementation of an automated learning center for the individualization of instruction

The special focus of the report is upon the adequacies and/or inadequacies of the evaluation sections associated with the proposals sampled. The frame of reference is that of Michael Scriven in his article "The Methodology of Evaluation" (AERA Monograph Series on Curriculum Evaluation, No. 1, 1967, pp. 39-89). Scriven offers an inventory of evaluation roles and procedures which can be used productively to assess the evaluation methodologies for the projects reviewed. Briefly summarizing Scriven's inventory, the following roles and evaluation procedures for the Title III projects were considered significant:

[*/] Prepared by Don D. Bushnell, vice president and director, Research and Development, Brooks Foundation, Santa Barbara, California.

Roles of Evaluation

Formative: Evaluation conducted during the developmental process for purposes of feedback to project directors and for subsequent revisions. Formative evaluation is synonymous with field sampling at critical junctures in the conduct of the project.

Summative: The final, overall evaluation of the results of the project. Concerned with outcomes and effects on project participants.

Procedures of Evaluation

Amateur Evaluation: The solicitation of reports from students and for members of community such as parents or the lay public, for judgment of the success of the project. In some instances, teachers may be asked to judge the success of a program, but will be judging in capacity as non-professionals.

Professional: Use of teacher as a subject matter expert to judge content or goals and equipment for classroom use. Hiring of research teams from university or expert consultants.

Process Research: All formative evaluation studies can be categorized as process research. Further, studies of teacher behavior in the classroom, or studies to discover the nature of a program through the testing of various hardware configurations can be considered process research.

Intrinsic Evaluation: Studies which assess the quality of the goals stated and concerned with the appropriateness of the curriculum goals to the selected student population. Some studies focusing upon intrinsic evaluation will be concerned with teacher attitude towards curriculum goals, test items, and worthwhileness of innovative procedures.

Pay-Off Evaluation: The examination of the effects of the programs on pupils, teachers, and other participants. Use of pre- and post-test control groups and split-pair analysis techniques.

Comparative: Studies which compare alternate programs and procedures with control groups, use standardized tests (participant against norms), as well as achievement and attitude scales.

Non-Comparative: Tests centered on the experimental programs or procedures alone.

In the following analyses, each of the major areas of computer application will be treated separately, i.e., data centers, media centers, and computer augmented instructional programs. The method of organization will be that suggested by Dr. Miller: (1) description of procedures and approaches to evaluation; (2) analysis of evaluation procedures; (3) recommendations for the future.

Computer Technology for the Augmentation of Curriculum

The major goal of introducing CAI into the curriculum in each of the four projects reviewed was to afford the pupil an opportunity to get a better and more thorough grasp of the mathematical base in the science subjects using modeling or stochastic procedures, and provide discipline in analyzing problems. Two used comparative studies for summative evaluation purposes. One considered establishing a control group of math students, but felt it unwise to exclude these students from a computer experience "just to obtain more data". A non-comparative evaluation with tests which emphasizes programming procedures and flow-charting will reveal the effects of the project, but the lack of comparative data doesn't tell us whether a less expensive or even a non-computer augmented approach could work as effectively.

In the Williamsport, Pennsylvania project which uses a more traditional form of PI for mathematics inservice education of elementary school teachers, summative or pay-off evaluation is of a comparative nature. The goal is to compare an inservice program offered via CAI with a program using conventional instruction. The major weakness of the study is that teachers using the computer facilities will also participate in a curriculum seminar taught by the project director while the control group will receive "conventional" content mathematics instruction and a curriculum seminar taught by the project co-director. The effects of the different instructional seminars will be measured on a special test developed at Penn State, an attitude scale measuring attitude toward mathematics, and teacher classroom behavior. The question remains, how do we know that differences, where they are relevant, are differences between the two seminar instructors or the use of CAI? The critical question of the effects of attitude changes and teacher behavioral changes on instruction is not measured, only assumed.

All projects proposing use of CAI in mathematics and the sciences do not evaluate results in terms of all objectives they proposed. They are particularly weak in the process analysis of hardware costs, configurations, and capacities. Little concern is shown for input and output procedures and the serious problems of student interfacing with the computer system. Three out of four of these proposals reflected a healthy concern for these practical problems in their objective statements. They spoke of establishing criteria to determine size, cost, and capabilities of the system under design. For example, the question was raised in one project as to the desirability of relative merits of giving students in math and physical science infrequent access to a full scale console with all its concomitant pinball effects or more frequent access to a time-shared teletype terminal, perhaps housed in their classroom. And in another proposal, the question was raised as to the desirability of owning or leasing a computer system versus the use of a service bureau system on a contract basis, or leasing of a WATTS line for long distance telephone connection to a large scale system at a nearby university. Turn around time (the time lag between checking out a program or problem solution on the computer and the next go-around after working out the bugs) and the computational powers of the system were also decision goals that were projected as major concerns in these projects.

While these objectives were appropriately underscored in the four proposals reviewed, no reference was made to them in the evaluation sections.

For projects involving considerable hardware expense, a concern for "process research", using Scriven's terms, is absolutely essential. Trade-offs on various equipment configurations should be a substantial part of the formative evaluation procedures for computer related projects. The inattention to goal congruency as represented in the four studies reviewed is some cause for concern. Apparently, the project initiators are capable of asking the right questions, but unconcerned about finding the answers.

Recommendations

1. FOR INNOVATIVE PROGRAMS APPLYING CAI TO CURRICULUM AUGMENTATION, CONTINUOUS STUDY OF ALTERNATE SYSTEM CONFIGURATIONS SHOULD BE MAINTAINED THROUGHOUT THE VARIOUS STAGES OF FORMATIVE EVALUATION.

The Automation of Media Centers

In the area of information retrieval or the automation of library procedures, application of new technology should be rationalized only on the basis of hard data demonstrating need. In the ___________ project to establish a regional computer-controlled media center (in fact an automated check-out system), there is no reference made to the need for automation because of the overload of a well used, manually operated distribution system. As a matter of contrast, it is explicitly stated that to evaluate the proposed system, it is necessary to initiate data collection procedures to ultimately get a measure on usage made by pupils, staff, and curriculum.

By constructing a computer-monitored information file, it is suggested the necessary data for evaluation will be provided. The first question is: why automation? Second: how will improvements with the new system be compared with the old? The evaluation section of the __________project contends that automation of a media inquiry system as proposed will automatically yield data on "changes in rate and scope of pupil achievement which result from exposure to various types of educational media." How does one measure changes without a preliminary data base? It will be difficult, too, to demonstrate that changes in pupil achievement are due to application of new media without (1) carefully controlling the media usages; (2) defining the behavior to be established; (3) stating the existing repertoire of skills the learner already possesses. To further state that the utilization of media will be correlated with a decline in school dropout rate is rather absurd.

The real benefits that may be accrued by automating a regional media center lie in the process of developing abstracts on media items contained in the centers. As proposed in the __________project, media records for the automated system will contain information on subject matter and indicate appropriate grade level or age range for study. This information will be prepared and placed in the file according to the KWIC (Key Word In Context) system. Reportedly, this process will facilitate the matching of key words which the teacher has typed into the system with the various media abstracts.

In this reviewer's opinion, if the abstracting of information is limited, i.e., to grade level and course, then the system doesn't need to be automated. From a monthly printed inventory list, the teacher could conduct her own random search for resources. But if, on the other hand, the item abstractions incorporate information that pertains to instructional strategy—for example, does the media form take an inductive or deductive approach, what examples are used, what concepts and related skills are taught, what are some of the documented results, et cetera—then the information load and retrieval potential may begin to dictate a need for automation. If, in addition, cross correlations between type of students, and previous test performance data taken by students of a similar population (tests designed to measure learning generated by specific media units) can be obtained, then automation and its concomitant costs for implementation are even more justified. In other words, the key to automation and a decision to move towards an automated system is based primarily on the information to be gained from the system and not on the rapidity of information retrieval which the system affords the user. Unless the need for fast retrieval of information has been well established through prior testing and data collection, this criterion alone is not sufficient cause for automation. Certainly a system of data retrieval via automated search and readout is not justified if the actual delivery of the information or media package being sought is handled by manual processes.

2. BEFORE AUTOMATION IS CONSIDERED A FEASIBLE MOVE, A DATA BASE CLEARLY SHOWING NEED AND APPROPRIATENESS OF APPLICATION SHOULD BE ESTABLISHED.

Automated Data Centers for Educational Decision-Making

The __________ IR system planners demonstrate in their statement of operational objectives a genuine understanding of a systems approach to automating information files. The design and implementation of the query language and hardware systems capabilities and the projected demonstrations on service bureau equipment prior to conversion to an "in-house" computer system are excellent procedures to follow in the implementation of new computerized systems. Unfortunately, in the evaluation phase, little, if any, insight or interest is shown in the adequacies of these procedures. The following is the sum total of the projected evaluation procedures for the __________ project:

> The field test version of the system will be evaluated by guidance, teaching and administrative personnel in the __________ Central Schools, and

> by the project staff, to determine whether the system meets its stated educational objectives. During development and implementation of the system, the EDP personnel will continuously evaluate sub-systems against criteria of size, cost, and efficiency of operation.

The question must be raised as to what specific criteria, size, cost, and efficiency of operation were included in the proposal. None were stated in the objectives section, therefore, it is literally impossible to judge the adequacy of the evaluation procedures. The generality of this evaluation statement makes it essentially meaningless.

INSERVICE TEACHER EDUCATION[*/]

This report is based on an examination of written proposals for 13 Title III projects, each of which has some major emphasis on the inservice education of teachers. The proposals examined were selected from a larger group of approximately 50 proposals, most of which also included a major emphasis on inservice teacher education. The 13 chosen represent a cross-section of the larger group according to the region of the country, submission date, type and size of school district (e.g., rural, suburban, urban; large, medium, small in population), and scope of project as indicated by the funds requested.

Proposal review was guided by 15 common questions or clusters of questions. The questions were phrased in order to provide useful information either to educators or evaluators in understanding the nature and adequacy of evaluation plans in Title III proposals.

Description, analysis, and interpretation of the evaluation plans of the 13 proposals studied are organized around these 15 questions. It is important to note that many proposals contained inadequate information for formulating clear answers to many of the questions.

Description, Analysis, Interpretation

I. Is there an evaluation plan based on an identifiable model or comprehensive strategy?

[*/] Prepared by Don Davies, executive secretary, National Commission on Teacher Education and Professional Standards, National Education Association.

Only four of the 13 proposals included anything that could be reasonably considered a model. Two of these four were informal and incomplete; the other two, comprehensive and specific. The most important fact to be noted here is the lack of attention to the need for a model or comprehensive strategy by the educators preparing the proposals. An enormous gap seems to exist between the views and knowledge about evaluation of these on-the-scene educators and the experts in evaluation who are writing papers and delivering speeches on the topic. The only evidence of communication between the educators and the evaluation experts in most cases is the adoption of some of the jargon of the latter by the former.

II. Is there provision for systematic gathering of information, including content, input, process, and product data?

Six proposals included provision for comprehensive data gathering. The others either did not indicate any provisions, or they stated that plans were going to be developed later. The need for gathering data of various kinds on a systematic and extensive basis was indicated in these six.

III. Are the data to be gathered quantitative or qualitative or both?

Six proposals specified both types. Three emphasized quantitative; two, qualitative; two simply didn't say. The reasons for selecting or emphasizing a quantitative or qualitative approach or both were cited in only two instances. The proposal writers were apparently not much concerned about this question.

IV. Is the approach judgmental or descriptive or both?

A direct answer to the question was given in only a few of the proposals. In eight, both judgmental and descriptive approaches were specified. One each would apparently be limited to one approach. In three cases an answer could not be determined.

No proposal included a rationale for emphasizing descriptive or judgmental data or both. The most salient point here is the lack of attention to the question.

V. Are behavioral objectives to be used in the evaluation process?

In seven proposals there were either direct or indirect specification of behavioral objectives as part of the evaluation process. In three, it was clear that behavioral objectives would not be a part of the process; in three others, no information was available. In the seven, only general recognition of the possibility of using behavioral objectives was provided. In no case, were the projects' objectives presented behaviorally in an adequate way.

In several proposals an effort to cast objectives behaviorally resulted in the identification of such non-behavioral general characteristics—"The teacher is a missionary." "The administrator has 'team spirit'." "The teacher is change-oriented."

The fact that behavioral objectives are "in" these days is reflected in many of the proposals. However, it is clear that the educators writing these proposals have an incomplete, inadequate, or distorted understanding of what behavioral objectives are and how they are related to measurement and evaluation.

VI. Are data being sought primarily about the process or the product?

Seven proposals included references to both product and process evaluation; three were limited to product data, one to process. Two others said nothing from which an answer to this question could be inferred. In the seven proposals the emphasis is on the side of product evaluation, usually data gathered by achievement tests. In most cases the product is defined as what children learn; in a few, however, the product is defined as changes in teacher behavior or changes in school program. There is considerable confusion about what is "product" or "process" and what roles each can play in evaluation.

VII. Is there an identifiable experimental design (formal or informal)? Does the proposal confuse research and evaluation?

Stufflebeam's contention that educators characteristically confuse evaluation and experimental research was not given strong support in the 13 proposals reviewed here. There was not one full-blown experimental research design. Five proposals included some form of experimental and control groups and pre- and post-tests. Only one proposal called for random assignment of pupils or teachers. One proposal specifically rejected an experimental design as inappropriate.

VIII. Are relevant and significant questions to be asked in the evaluation process identified? Is determining the relationship between stated objectives and possible outcomes indicated?

In every proposal some effort was made to identify relevant and significant questions. In the majority of cases this was done in a vague, general, and brief way. In only two or three proposals were the power and clarity of the questions adequate to provide guidance to the evaluation process.

All but three of the proposals recognized the importance of the relationship between stated objectives and possible outcomes in the evaluation process.

IX. Is there provision for feedback from the evaluation process to aid program improvement; to aid decision-making?

Seven proposals mentioned feedback for program improvement; six mentioned feedback for decision-making purposes. All the others said nothing on the topic or were unclear.

With two or three exceptions, little emphasis was given to either of these important purposes of evaluation; program improvement or more effective educational decision-making. However, three of the proposals which seemed strongest in general gave important attention to the uses of evaluation. The material in these proposals, taken as a whole, could be used to support the contention that evaluation is something that one must do but not something that is terribly relevant to the real world in which educational decisions are made.

X. Is there provision for obtaining the perceptions and opinions of key groups involved in the project (e.g., pupils, parents, teachers, administrators, subject-matter specialists, the public) as part of the evaluation plan?

Six of the proposals made such provision; seven did not. Three proposals call for extensive gathering and utilization of the perceptions of the participants and others, including elementary or secondary school pupils and parents. Two proposals provided machinery for continuous participant reaction and feedback.

No sophistication about evaluation is required to recognize the importance of the perceptions of participants. Hence, it is surprising that more than half of the proposals failed to mention such perceptions as data.

A paradox: One proposal which includes an administrator training component and a teacher training component, calls for heavy participant reaction and self-evaluation by administrators but none by teachers. The administrators are to be asked to evaluate the program on the basis of how they feel; the teachers are to be evaluated by outside experts.

XI. How strong is the commitment to local option (e.g., evaluating in terms of local purposes, conditions, and opinions)? Are data to be cast against identified standards, either relative or absolute?

In only five instances was there any evidence on the local orientation question. In these five, three evaluation plans were strongly oriented locally; two indicated a deliberately broader context.

In view of the usual strong support for local control among American educators, it is surprising that more did not specify that the evaluation plan be closely oriented to local purposes, needs, problems, and conditions.

The word "standards" was almost never used. There was not a single proposal which clearly stated that data gathered would be examined against any existing standards, relative or absolute. Standards were simply ignored, which must be taken as confirmation of the aversion of educator to applying external yardsticks to local data. An outside observer would certainly find astonishing this ignoring of standards.

XII. Does evaluation seem to be a genuine and intrinsic part of the proposed project as contrasted to something that is included in order to meet a requirement?

A firm and reliable answer to this question is not possible. Reading written proposals provides shaky basis for making judgments about motivations and attitudes. Some light on the question is provided by the fact that in eight proposals there is no reference to evaluation except in the evaluation section. In the other five, references to evaluation are found in appropriate places throughout the proposal.

Four proposals give major emphasis to evaluation; one gives moderate emphasis; the others give only minimal emphasis.

XIII. Are staffing and budgetary provisions for evaluation included?

What individual or group is responsible for the evaluation?

The most direct evidence about the emphasis given to evaluation lies in the fact that six of the 13 proposals make no specific mention of staffing for evaluation. Five of the proposals assign responsibility for evaluation to the project director; four to an outside consultant or agency; two to a special staff member for research and/or evaluation; one to the chairman of an advisory committee. Two proposals do not assign the responsibility at all.

Only one of the proposals gives any details about staffing plans and the qualifications of those who are responsible for evaluation. Only one of the proposals makes any special point of supporting the need for funds to conduct an adequate evaluation.

Evaluation does not appear as a separate budget item in any proposal. Money for purchase of tests is included in all of the proposals which call for the use of standardized instruments.

Incomplete evidence about staffing and funds for evaluation must be taken to mean a lack of importance assigned to evaluation by those who prepared the proposals.

With four or five exceptions there is no recognition that special expertise (beyond the normal general qualifications of educational administrators) is required for those who will plan and conduct evaluation programs.

XIV. What data gathering techniques are identified? What specific instruments are mentioned? Are samples of instruments to be used included in the proposals?

The techniques, with frequency of mention were as follows:

achievement tests-pupils, standardized	7
achievement tests-pupils, teacher-made	2
achievement tests-teachers, standardized	1
pupil attitude/personality tests, standardized	2
teacher attitude/personality, standardized	3
teacher attitude/personality, locally-made	2
test of critical thinking, pupils	1
interview	2
case studies, pupil profiles	3
pupil diaries	1
teacher diaries	1
critical incidents (supplied by teachers)	1
pupil grades	1
essays, teacher and administrator	1
questionnaire, parents	2
questionnaire, teachers, participants	3
pupil rating of teachers	1
teacher rating of pupils	1
parent self-evaluation	1
teacher/administrator self-evaluation	1
classroom observation:	
by peers, administrators	5
by trained observers	3
by video-tape	1
using interaction analysis	1
evaluation of teacher lesson plans	1
staff meetings, workshops, conferences	2
review by advisory committee	1
review by curriculum specialists	1
survey of school activities	1
quality and nature of student extra-curricular activity, attendance	1
teacher turnover	1

Only five proposals named any specific instruments. The following were named: Iowa Basic Skills Test (2); Iowa Test of

Educational Development; Metropolitan Achievement Test; California Achievement Test; Stanford Achievement Test; Survey of Reading Practices in Georgia Schools; Teaching Learning Process Analysis Inventory; Holpines Climate Index; Flanders Interaction Analysis; Watson-Glazer Test of Critical Thinking.

Only two proposals included a copy of an instrument to be used. One of these was a locally-made teacher opinionnaire about a model school; the other was a locally-made survey of teaching practices.

Only two proposals mentioned that instruments were to be field tested before used in the evaluation program.

The number and variety of specific instruments named were smaller than might have been expected in light of the vast number of instruments available. It was also surprising that so little attention was given to the need for field testing.

XV. Is the proposed evaluation plan innovative or traditional?

Only two of the 13 proposals could be said to include an innovative approach to evaluation. It is extraordinary that proposals for innovative programs could include so little effort to innovate in evaluation approaches. The most likely reasons for this circumstance are lack of interest, expertness, and confidence with relation to evaluation.

Recommendations

1. THE OFFICE OF EDUCATION'S REQUIREMENTS AND EXPECTATIONS FOR THE EVALUATION COMPONENT

OF TITLE III PROJECTS SHOULD BE RAISED SIGNIFICANTLY.

No proposal should be accepted unless it includes a comprehensive plan for evaluation. The evaluation plan should include all of the following:

(a) A clear statement of the rationale for the evaluation approach to be followed.

(b) Specific indication of adequate staffing and budgeting for evaluation, including a listing of the qualifications of those responsible.

(c) A clear indication of the kinds of data to be gathered and the instruments to be used.

(d) A clear indication of plans for utilization of what is learned from evaluation.

The material on evaluation in the latest version of the Title III guidelines should be redone. It should be more specific and more detailed.

2. TITLE III FUNDS SHOULD BE ALLOCATED TO PROVIDE TO LOCAL SCHOOL DISTRICTS CONSULTANT HELP BY EVALUATION EXPERTS (OR AS AN ALTERNATIVE POSSIBILITY, STATE DEPARTMENTS OF EDUCATION SHOULD BE PROVIDED FUNDS TO USE FOR THIS PURPOSE).

Those preparing Title III proposals apparently lack sophistication and expertness in evaluation. They are educators not evaluation experts. If evaluation components of PACE projects are to be improved significantly, outside expert assistance will be needed. Exhorting local school district officials to "do better on evaluation" will not do much good, unless specific ways are found to provide expert knowledge and skill which is now lacking.

In addition, brief papers about evaluation by people such as Stufflebeam and Stake could be provided, along with descriptions of the most effective evaluation programs in operation, to local school districts preparing Title III proposals.

3. EMPHASIS SHOULD BE GIVEN TO USES OF EVALUATION AND SHOULD BE PROVIDED BY TITLE III OFFICIALS—THROUGH CONFERENCES, WRITTEN MATERIALS, AND FILMS.

One of the reasons so many school officials give so little emphasis to evaluation is that they do not see the potential constructive uses of the result of evaluation. Educators may need specific help in seeing how the results of evaluation can help them function more efficiently and effectively.

SPECIAL EDUCATION[*/]

Probably in no other area of education is there a greater need to devise, try out, and test improved procedures for the education of handicapped children. Educators generally are inclined "to get on band wagons"—chasing after promising leads or arm-chair ideas. Special educators, understandably, may be particularly prone to this since their role, status, and efficacy are being seriously questioned. For this reason, Title III has great promise. It can enable special educators to test out procedures designed especially for a newly enriched, flexible, and effective regular school program. It is paramount that these innovations be carefully evaluated before their wide-spread usage is advocated. With modern school changes, especially the extra services provided by compensatory education, the challenges are two-fold: to determine if pupils formerly classified as handicapped can now make satisfactory progress in diversified and enriched regular school programs, and to devise effective special education services in local schools for children with more extreme and complicated forms of learning disabilities.

Evaluative Aspects of Selected Title III Projects

A most cursory examination of proposals for Title III grants submitted for funding in 1965 and 1966 reveals that essentially no

---

[*/] Prepared by Lloyd M. Dunn, director, Institute on Mental Retardation and Intellectual Development, and professor of special education, George Peabody College for Teachers.

attention was given to evaluation. With the advent of the Manual for Project Applications and Grantees in mid 1966, requiring a section on evaluation (see p. 48 of the May '67 revision of the manual), and with a Chapter V outlining the evaluation requirements, the attention paid to this topic changed dramatically. While one can question the rigorousness of both the stimulus and the response, it is clearly evident that setting expectancies on the part of the grant-awarding agency did elicit attention and concern for this matter on the part of the grant-seeking systems.

It would seem that little would be gained in taking a historical look at the essentially non-existent evaluative techniques outlined in the proposals submitted prior to 1967. Instead, it seemed more fru ful to sample a cross section of proposals submitted under the new grants manual. To accomplish this, as seen below, all applications were classified and a representative sample of 12 of the 1967 proposals selected to give a cross section of the types of evaluative procedures now being proposed. This should better enable one to speak to the question: What is the quality of the evaluation process now being proposed, and how can it be improved?

More requests were submitted for Diagnostic and Learning Laboratories than any other type of program. Of the 19 received, three were examined in detail. Even the weakest suggests a disaffection with present and past diagnostic centers which have been run by

Area or Type of Proposal	Proposals received	Prior to 1967 proposals	1967 proposals	1967 sample examined
1. Diagnostic and learning laboratories	19	13	6	3
2. Programs for emotionally disturbed/behavior problems/brain injured children	6	3	3	3
3. Programs for mentally retarded/slow learning children	16	9	7	3
4. Comprehensive special education programs	5	2	3	3
5. Programs for the cerebral palsied	1	-	-	-
Totals	47	27	19	12

An outline follows of the types of evaluation methods proposed in the sample of 12 projects examined. (see page 43).

---

psychologists, social workers, and physicians. They make educational diagnoses, label children with specific physical and psychological disabilities, and provide little or no help to the teacher in devising an adequate educational program. The trend is clearly to leave educational diagnoses largely to educators and to broaden the centers from diagnosis to the development of instructional procedures tested and found effective for the child.

The three most common forms of evaluation mentioned were

Evaluations Methods Proposed	Totals	Diagnostic and learning laboratories	Programs for the emotionally disturbed and brain injured	Programs for the mentally retarded and slow learning	Comprehensive special education programs
1. Pupil progress	22	7	4	9	2
- standardized tests	7	2	0	3	2
- non-standardized tests	2	1	1	0	0
- pre- and post-tests	3	2	0	1	0
- follow-ups	5	1	1	3	0
- case studies	5	1	2	2	0
2. Pupil reactions	2	0	1	1	0
- interview/conference	1	0	0	1	0
- questionnaire	1	0	1	0	0
3. Teacher reactions	6	2	1	2	1
- case studies	1	0	0	1	0
-interview/conference	3	1	1	1	0
-questionnaire	2	1	0	0	1
4. Director evaluation	2	0	0	1	1
- case studies	0	0	0	0	0
- interview/conference	0	0	0	0	0
- questionnaire	0	0	0	0	0
- reports	2	0	0	1	1
5. Outside consultant evaluation (incl. contracts with reg. labs & universities)	9	4	2	1	2
6. State department personnel evaluation (including PACE coordinator)	1	0	0	0	1

Evaluations Methods Proposed	Totals	Diagnostic and learning laboratories	Programs for the emotionally disturbed and brain injured	Programs for the mentally retarded and slow learning	Comprehensive special education programs
7. Parents' reactions	2	0	0	1	1
8. School administration reactions	3	0	2	1	0
9. Statistical reports (persons participating, materials used, etc.)	4	3	0	0	1
10. Para-education personnel reactions (psychologists, etc.)	1	0	1	0	0
11. Comparative studies (research involving contrast groups)	3	0	1	1	1

the measurement of pupil progress, outside consultant evaluation, and statistical reports on the usage of the facilities. In not one case was an alternative to the clinic proposed. In short, no contrast procedures are being compared against this new-type, diagnostic and learning laboratory. Clearly, therefore, the evaluation will net little or nothing of value in testing the efficacy of these centers. All will accumulate data to support them, but this will be largely busy work performed as a requirement to secure the federal money. What is needed is to compare their effectiveness against alternate procedures. For example, one could contrast the centers with the effects of assigning an educational diagnostician and therapist (clinical teacher) full-time to each school, rather than centralize all of them in a clinic. Until alternate procedures are compared against these centers, we will have essentially no evidence to support their continuance. If this cannot be seen, there is little hope for educational evaluation.

Three of the six requests were examined in detail for programs to serve emotionally disturbed/brain injured/perceptually impaired children. Labels for handicapped children are in a crucial state of flux largely because medical advances, drug and other ingestions, etc., are presenting the schools with more and more multiply-handicapped children with a variety of bio-behavioral disorders, some of which are chronic and others of which are developmental in nature. It is small wonder that these children who present a multiplicity of overlapping problems to the schools, society, and parents are variously labeled, emotionally disturbed, behavior problems, learning disorders, perceptually impaired, and motor handicapped. These children present a critical and emerging challenge to education. Thus it is good to see proposals submitted in this area. Too, it is not surprising that these requests are variously labeled. As in case of the clinics, the evaluation emphasis was on measures of pupil gain and views of outside consultants. More attention was given to reactions of school administrators and others. Too, one project indicated it planned comparative studies though no specifics were given.

Three of the 16 proposals for programs to serve mentally retarded and slow learning children were examined. Here was the area of second greatest emphasis after clinics. Generally, the proposals are to establish schools for educable mentally retarded adolescents, and for outdoor living programs. Here, almost all the evaluative emphasis was on the measurement of pupil growth. One planning grant was asking some very pertinent questions; namely: Will pupils make more progress in a special class or a regular school program? Will there be fewer dropouts, less disciplinary problems in school, and less delinquent behavior outside the school as a result of this special program? Will pupils in the special program become accepted members of the school society? Whether research procedures will be devised during the planning stage to study these questions remains to be seen. It will take a good deal of sophistication to do so.

The more rural, sparsely populated areas tended to be the ones submitting proposals for comprehensive special education programs. Of the five submitted, three were studied. Generally, there were plans to develop programs which would cut across "categorical" (specific disability label) lines. Most intended to collaborate with a regional college or university. Generally, diagnostic and therapeutic centers were proposed as part of the package. Evaluation of these programs was essentially non-existent. Clearly the applicants were among the least sophisticated, the communities in the greatest need, and the likelihood of innovative procedures least. There was one exception where seven modes of evaluation were proposed: (1) statistical reports (cataloging), (2) visitor questionnaires, (3) evaluation by State department of education personnel, (4) evaluation by university personnel, (5) co-director's reports, (6) research consultants, and (7) comparative studies (which are not clearly outlined). Very probably, a behavioral-science type, special educator in a college had played a major role in constructing this application. Except for this one request, it will be difficult to determine whether these projects are successes or failures. Yet, one of the greatest needs is to extend special day and boarding school programs to rural children with learning disabilities.

How, then, does one stimulate greater quality in the evaluation of projects for handicapped children funded by Title III? Does the CIPP

Model provide a useful procedure? This question will be examined next.

The CIPP Model and Special Education

The CIPP Model, it seems to me, is an extremely useful strategy system for the project applicant in initiating, developing, and monitoring (not just evaluating) programs. Unfortunately its author has chosen to set up a straw man and then attack it. In so doing he has negated much of the cogency of his argument. He has thrown the baby out with the wash by disparaging the culminating type of evaluation, namely product evalution. In fact, he downgrades product evaluation so badly that he cannot adequately defend it later. He has chosen to define evaluation so broadly that it lacks utility for answering the crucial question as to whether the innovative program is an effective one.

Context study is useful to the project applicant as he conceptualizes an innovative school program. One must study community and school strengths and weaknesses, and interface them against pupil characteristics and unmet needs to identify problems, hypotheses, goals, or objectives.

Input study is important to the project applicant, as he assesses the resources needed to attack the problem, as he reviews the literature on what has been done under similar conditions, and as he devises the creative and innovative strategies (courses of action) to attack the problem.

Process study is highly desirable (if not crucial) as the project director pilots, monitors and improves upon his educational intervention. Few would advocate, while a procedure is being developed, the employment of "an ivory tower specialist" to set down a hard and fast experimental design with experimental and contrast groups, with randomly-assigned students, with treatment procedures held constant, and with pre- and post-tests. This should come only after the crudely conceived innovative program has been perfected to the point where it can be considered worthy of a trial. Who would argue against planning and perfecting grants to field test, modify, and improve on an innovative idea?

But all of these three strategies—the so-called context, input, and process evaluation (the CIP of CIPP)—should be funded by planning grants, should not be considered evaluation *per se*, and should precede the operational grant.

Eventually the taxpayer, his representatives, the public, and the teaching profession have a right to know whether this carefully-developed innovation is more, less, or equally as effective as conventional or alternate procedures. This demands the use of the most sophisticated research designs, measurement, instruments, and inferential statistics available to behavioral scientists. Surely the best procedures available to behavioral scientists are not too precise or too good to apply to programs designed to shape the behavior of our Nation's youth. Without some frame of reference—-contrast treatment groups, national norms, etc., —there is simply no valid way to know whether to continue, terminate, modify, or replicate the program. In short, the crucial evaluation is one which measures the effectiveness of the program—even though the need is also great and necessary for the accumulation and thoughtful analysis of relevant information (data) for identifying the problem, for devising strategies, and for improving on the initial conceptions.

It is therefore recommended that the CIPP Model be refined and used by Title III to provide a series of logical strategies for project personnel to follow as they identify, develop, and test an innovative program. However, only the last quarter of it (namely the product evaluation, the past P in CIPP) deserves the label evaluation (which is defined in the dictionary as "determining the worth of"). Common sense suggests that evaluation in our context, means just that—to evaluate the worth of an innovative program.

A number of educators pervert the accepted societal meaning of evaluation when they define it so broadly that it is in fact a strategy of data collection and study to conceptualize and improve upon an idea. Nothing short of a hard-nosed and critical look at these innovations—utilizing comparative data, valid and reliable measures, and sophisticated means of data analysis—will suffice to prevent the Nation from being lead down the garden path, strewn with flowery generalizations based on soft data about the effectiveness of our innovations. Title III could be scuttled, education set back a decade or more, and legislators thoroughly dissaffected by utilizing the CIPP Model— in its present form which neglects and disparages product research— in the Title III grants manual as a model for evaluating the effectiveness of innovative programs. However, if adequately modified and a balanced case made for product evaluation, there is no reason why the CIPP Model would not be most useful in the manual as an overall strategy for project personnel to follow as they slowly and methodically devise, develop, refine, and test their innovations.

Recommendations

From a study of the applications, it would appear that Title III has been a crash program. In general, the applications were written too quickly and padded too voluminously, by applicants who had too little training as applied behavioral scientists and too little knowledge in special education, and who generated too few creative ideas. What is needed now is a thorough re-structuring of the Title III grant procedures to build in greater assurance that projects funded in the future will be more carefully formulated, will reflect the best thinking of the Nation's authorities in the field, will have been gradually developed, modified, shaped, and perfected, will (when ready) be carefully evaluated, and (when demonstrated as effective) be replicated. It is suggested that this can best be accomplished by moving project directors through four phases (hopefully funded separately as each phase is successfully completed).

Phase 1: Innovative Program Planning Grant (1 to 2 years)

Here the manual should outline minimum requirements for a very brief application to secure a mini-grant or contract of probably a flat $25,000 a year for one to two years. With this support and during this phase, the manual should require the applicant to go through the content and input strategies in the CIPP Model. Here the applicant would be asked to involve as consultants the most creative and divergent thinkers in the Nation who are knowledgeable about the problem to be attacked. (No outside evaluators and critical thinkers would be brought in to stifle creativity.) The goal would be to formulate an innovative program in sufficient detail to write a proposal to field test and refine it.

Phase 2: Innovative Program Refinement Grant (up to 3 years)

Here the manual should require the applicant to build in during this phase of grant support, strategies to accomplish the process study outlined in the CIPP Model. During this phase, the local school system would usually operate a pilot program to field test the innovation. The emphasis would be to develop, monitor, refine, and modify the procedures to be followed. Such techniques as freeing up educators to detail instructional programs, and test and modify them, would be encouraged. Here the consultants should be trouble shooters/problem solvers of the clinical types (master teachers, consultants, supervisors) who can study and monitor the development of the program, and suggest ways for improving on it. This phase could last up to three years and perhaps even longer. It might be amalgamated with a Phase 1 proposal, but this is not recommended.

Phase 3: Innovative Program Evaluation (operational) Grant (3 or more years)

Innovation without evaluation is dangerous, if not foolhardy. Thus, the manual would emphasize during this phase of grant support, the need for product evaluation with the most sophisticated designs available. At this point, it is recommended that behavioral-science type, critical thinkers and evaluators be brought in. Here there seems no other way, for some years to come, except for local school systems to contract with colleges, universities, regional laboratories, etc., to conduct these evaluations. There is especially a great paucity of research-oriented scholars in special education. What few there are, are largely in college and universities. Practitioners must look to these persons (and to others in statistics, measurement, and research design) for a critical evaluation of their innovations. Adequate research design, subject selection, measurement instruments, and inferential statistics are generally beyond the training of special educators in local school systems. True, there is a shortage of sophisticated researchers in all aspects of education in local school systems. It is just that this is especially so in special education.

There is a natural temptation for innovators to be enamoured with their ideas. Thus, there would be an understandable temptation for project directors to gather the types of data which are most likely to be favorable. Here they could give pre- and post-tests to measure pupil growth—but time and almost any intervention will produce some positive gains, but would they be significant? And of what value will be the opinions gathered by the director and his staff? There is a tendency for people to be kindly, charitable, and favorable about another person's brain child. But is such evidence valid? Not very likely. What is needed is an external evaluation of the effectiveness of a program. This can range from comparative studies by non-involved researchers, to evaluations by uninvolved experts who bring to the project as site visitors, hopefully, a general frame of reference and a critical, knowledgeable eye, acquired in developing their expertise over the years. Looking at the relative effectiveness of a variety of strategies built into the program, examining which pupils fail and succeed in the project, etc., present helpful alternatives for evaluating programs.

The biggest decision of all for the granting agency is whether or not to terminate or replicate the innovative strategy on a national basis. To help in this decision is why the government is justified in investing the taxpayers' money in Title III projects. And decisions become important in the context of practical choices, and evaluative data which suggests which is best.

Phase 4: Exemplary Program Support Grant (indefinite)

It is inadvisable to fund exemplary programs until their efficacy has been established. Thus, Title III funds should not be used to support such programs for another three to eight years, or until the innovative projects have moved through Phases 1 to 3 above. Who knows how to do it right, until it has been demonstrated to be more effective than alternate procedures?

However, eventually this Phase 4 could become the crucial one for implementing changes in American education on a broad basis. Thus, there would seem to be merit in funding exemplary programs which have been tried, tested, and found effective over a number of years.

Here the outside visitors would be still another sort. They would tend to be the leaders in special and general education in Federal, state, and local school systems in positions to influence changes in their schools. They can stimulate replications of exemplary programs on a broad basis. Any talk of dissemination in the first three phases—except to inform the field of what is being attempted so as to prevent unknowledgeable duplication—would be premature.

THE CULTURAL ARTS*/

Forty proposals in the cultural arts were sent for review and of these, twenty were read with respect to the evaluation procedures employed. The following comments describe my views regarding their adequacy.

It should be made clear at the outset that the evaluation of learning in the cultural arts is a task beset with a variety of special difficulties. Perhaps the most significant of these difficulties is the fact that traditionally those working in the cultural arts—the visual arts, music, literature, poetry and drama—have not placed much faith in objective assessment of artistic learning as a means of judging the effectiveness of their programs. For many, the type of learning or growth that emanates from artistic experience is both subtle and idiosyncratic. Hence objective measures of standard outcomes are often viewed as trivial compared to what is considered more important and fundamental in aesthetic learning.

This orientation as well as other factors have created a virtual vacuum in the area of fine arts tests. Of the 2300 tests listed in Tests in Print, only 14 are in the fine arts and of these many are out-of-date and inadequately standardized. Thus there are few evaluation tools available in the fine arts compared to the number available in other fields and certainly fewer useful for assessing the consequences of unique programs. Thus the context in which evaluations in the cultural arts reside is one that tends to have little disposition toward objective evaluation and few instruments appropriate for evaluation when unique objectives are formulated.

---

*/ Prepared by Elliot W. Eisner, associate professor of education and art, School of Education, Stanford University.

It is not surprising, therefore, that the proposals in the cultural arts should be weak in the area of evaluation. These weaknesses stem, in part, from the conditions I have described as well as from a variety of other factors.

Many of the proposals I reviewed, indeed the majority, were designed to provide short-term supplementary activities to children in the field of the cultural arts. These activities include trips to museums and plays, attendance at concerts and other types of activities through which cultural exposure can occur. While these activities are frequently supported by instructional material that teachers can use to prepare children for the cultural event, one gets the impression that the programs are superficial and peripheral in nature. The fact that such programs tend to be short term makes the likelihood of significant learning remote. I believe proposal writers intuitively recognize this and hence tend not to specify anticipated behavioral change as a result of such participation.

A second characteristic of the evaluation sections of the proposals in the cultural arts is a general lack of attention to the various aspects and populations of the program that can be evaluated. For example, it is the rare proposal that attempts to evaluate the reactions of various populations such as pupils, teachers, administrators or parents to aspects of the proposed program. When evaluation procedures are described, they almost always deal exclusively with student behavior.

This lack of analysis of the potential candidates for evaluation is only symptomatic of a general lack of evaluation competency. Given the dearth of competent people in the field of educational evaluation this lack of attention to the assessment of different populations and the relationship among data secured from these populations is not surprising.

About 30 to 40 percent of the proposals reviewed indicate merely that evaluation will occur and that appropriate evaluation instruments will be obtained or constructed but little more data than that is provided. When this strategy is taken to describe evaluation it is impossible to obtain an understanding of what will occur.

A third characteristic of the proposals in the cultural arts is their general vagueness with respect to educational objectives. As indicated earlier, the assessment of artistic learning is especially difficult as is the specification of objectives that are not trivial. Nevertheless there are some outcomes of cultural art experience that can be identified at reasonable levels of specificity and which are likely to be appropriate for all or almost all students. In general, however, these generic and specifiable objectives have not been described; hence the generality of objectives makes evaluation difficult.

Another characteristic of the proposals on the cultural arts is a tendency to employ published but inappropriate instruments for purposes of evaluation. One gets the impression that such instruments are being used because they can be purchased and not because they are appropriate or an important part of the project. The use of available but inappropriate assessment tools will tend to militate against the identification of behavioral change, especially when the goals of the program are innovative. The majority of the proposals make no indication of the particular types of instrumentation to be used and in no proposal did I find a set of instruments to be used.

Another characteristic of the evaluation sections of the proposals is a complete lack of attention to the description and evaluation of processes of the program. Since many of the activities suggested by the proposals are large scale, short-term activities, evaluation of the processes through which the activities proceed is difficult to obtain; and if obtained, might not be reliable indicators of what was occurring. But if the procedures used in a really innovative and effective educational program in the cultural arts is to be useful to other school districts some careful description of the program and the strategic supports necessary for its success need to be provided. A careful description of "input" as well as "output" is necessary if dissemination is to be effective.

The general impression that I obtained after reviewing the proposals was that proposal writers were naive about evaluation methodology. It has been recognized by those of us working with Title III projects that evaluation of on-going innovative programs is an exceedingly complex task. The classical pre- and post-test experimental paradigm is inadequate and inappropriate for providing dynamic feedback to innovative programs. Given the lack of sophistication in the field of education regarding the subtleties of curriculum evaluation it is unreasonable to expect those in the field to possess evaluation expertise. Judging from the proposals in the cultural arts, they don't! The evaluation sections are weak, non-analytic and frequently use inappropriate instruments to assess poorly specified outcomes. Many of the proposals merely nod in the direction of evaluation by indicating that it will be employed or that consultants will be obtained but for the most part evaluation tends to be an "add-on".

It seems to me that it might be of use to proposal writers to have a document prepared by the Title III office that would describe in lucid, non-technical terms some of the approaches and procedures that might be employed in an evaluation program. Such a document might facilitate more adequate evaluation procedures. The evaluation sections in the cultural arts proposals are so skimpy that almost any effort to make them more rigorous and thoughtful would be an improvement. It might also be well if universities would sponsor evaluation institutes for Title III personnel and for school districts anticipating preparing a Title III proposal.

SOCIAL STUDIES[*/]

Provision for evaluation has been one of the weakest part of proposals for Title III projects, according to studies that have been made of successive groups of proposals received during the first two years ot Title III operations. The proposals which were rated in these studies were not restricted to social studies, but dealt with various aspects of the school program. Some improvement in planning for evaluation was apparent when second-year proposals were compared with those received during the first year but, when the projects were rated on 15 criteria, second-year proposals were rated lowest on the criterion, "provision for evaluation of the project."

To what extent do projects dealing with social studies share this weakness which has been identified in the total group of proposals? This examination of a sample of proposals for social studies projects cannot provide a definitive answer, but does give some evidence on the point.

The Sample

Fourteen Title III projects dealing with the social studies area of the school curriculum constitute the sample on which this report is based. This sampling was selected at random from the Title III projects that have been funded in the field of social studies. Three are

---

[*/] Prepared by Dorothy M. Fraser, coordinator of social science, College of Education, Hunter College of the City University of New York.

from the submission period ending May, 1966 and eleven are from the period ending January, 1967. The size of grants involved ranges from $8, 502, a planning grant, to $117, 040, for the first year of a project that includes planning, pilot, and operational activities. Two of the three from the earlier submission period are planning grants, while the third is an operational program to implement an earlier planning grant. Of the 11 submitted in January, 1967, only two are devoted exclusively to planning; two are operational programs and the other nine involve planning plus pilot activities, or planning, pilot, and operational activities.

The 14 projects vary widely in scope and emphasis. Four are concerned with establishing regional centers to provide stimulation and services for local curriculum planning and/or demonstrations to encourage dissemination of innovative curriculum and instructional practices. Two involve the development of centers to facilitate use of local or regional resources in school programs. Three are devoted to the development of a specific social studies course and two others focus on activities and materials for the enrichment of a specific course. Planning and implementing a new social studies curriculum is undertaken in five of the projects. One of these deals with the primary grades, with anticipation of extending the project to the upper elementary grades, while in the other four planning is on a K-12 basis. (Two of the K-12 projects are being conducted by regional centers referred to above.)

Major directions of change that have been much discussed by social studies specialists in the past decade are reflected in this sampling of Title III projects. The use of multi-media materials, including out-of-school use of local and regional resources as well as a variety of learning materials to be used in the classroom, receives major emphasis in half of the projects. Cooperative curriculum development, bringing together the staffs of local schools in a region and/or drawing extensively on academic and professional personnel of colleges and universities in the area, is a basic factor in two-thirds of the projects. Teacher participation and inservice programs for teachers are emphasized in many of them. Approaches stressing inquiry and the development of important concepts and generalizations by pupils are found frequently.

Plans for Evaluation

Of the 14 proposals reviewed, two give no plans for evaluating the results of the project but indicate that planning for evaluation will be undertaken. Both are applications for planning grants, which are not required by Title III guidelines to include provisions for evaluation. Each of the other 12 proposals, two of which are for planning grants only, give at least some indication of plans for evaluation. This attention ranges from broad statements, which indicate some types of procedures to be used or state that comparisons of experimental and control groups will be made, to (in one case) a listing of general procedures to be used in checking progress toward each goal of the project. Titles of standardized tests to be used are given in two of the proposals.

In each of the 12 projects, some measures of pupil performance are to be used in evaluating the program. Comparison of achievement of experimental and control groups is the most frequently cited procedure, found in seven of the 12. Pre- and post-testing is mentioned in four programs; in three of these the target population apparently is to include only pupils in project classes. Standardized achievement tests are to be used in six projects, and teacher-made achievement tests in two others. Whether or not there is increased participation by pupils is to be studied in six programs by collecting evidence on such points as: use of library, study centers, and workrooms; class attendance; participation in extra-curricular activities related to social studies; and socio-civic action such as volunteering to assist in anti-poverty programs. Other approaches for evaluating pupil performance that are mentioned in one or more proposals are: observations by teachers, project staff, supervisors; completion of attitude scales by pupils; pupil self-evaluation; analysis of pupil products and of recordings and pictures of classroom sessions.

Evaluation of teacher performance as a measure of the project's success is indicated in five of the proposals, four of which mention only one approach (but not the same in each case). Observation by supervisors or project personnel is the approach most frequently relied upon.

In most of the projects, evaluative reactions are to be obtained from more than one sector of the personnel that are directly or indirectly involved. Teacher reactions are to be considered in seven of the programs, outside consultants are to evaluate in five. The project director and staff will give evaluative reactions in four projects, as will administrators and supervisors. Pupil reactions will be obtained in three projects, parent reactions in two, and the wider community (civic and business leaders, etc.) will be asked for reactions in one.

The instruments to be used in obtaining these evaluative reactions include: questionnaires and opinionnaires (4); rating scales (4); check lists (3); anecdotal records (3); interviews (2); attitude scales (1); sociograms (1); interest inventories (1); and open-ended questions (1). In most of the projects from three to five of these instruments are listed.

Naming the evaluative procedures listed in the proposals indicates one aspect of the provisions for evaluation. Other factors must also be considered in examining the adequacy of the assessment plans.

How fully are the plans for evaluation spelled out? The space devoted to evaluation in the 12 proposals that dealt with the topic ranges from nine lines to three single-spaced pages. In most of the proposals the discussion is extremely general and brief (about one double-spaced page). Usually it is stated that procedures and instruments must be developed because adequate ones are not available, but little or no indication is given about how this job will be tackled except that outside consultants would do it or provide guidance. While the length of a statement is not a test of its quality, the more extended discussion of plans for evaluation found in three of the proposals were the ones in which the authors displayed some familiarity with recent literature on evaluation in social studies and in which some approaches to the development of procedures and instruments were at least partially spelled out. One of the more fully developed sections on evaluation cites existing instruments that may serve as a starting point for construction of suitable checklists, scales, and questionnaires. In another proposal that has a fuller treatment of evaluation, types of data that will be collected to feed into the construction of instruments are listed.

Who is to formulate evaluation instruments and approaches for applying them? In about half the proposals this question is not directly considered. In most of the rest, the job is apparently to be done largely by outside evaluators or evaluation consultants. In one, however, it is stated that participating teachers and administrators are to work with the evaluation consultant in deciding on approaches and constructing the instruments. The rationale given for involving teachers and administrators in this way is that they can contribute to the process, that they will gain new insights as they do so, and that they will be more supportive of—or at least less threatened by—the evaluation of the project if they have helped to develop the assessment program.

Are the evaluative measures related to project objectives? In general, yes. But in nine of the 12 proposals there is little or no effort to relate a particular project objective to particular evaluation procedures. The objectives themselves are usually stated at a general level without defining their components or stating them in terms of specific behaviors to be anticipated. This lack of clear definition of goals at a less-than-global level is probably a factor in the failure to plan specifically for measuring progress toward them.

Is the need for funds to support evaluation taken into account? Five of the 12 proposals have no budget item for evaluation. Six contain items for services of outside evaluators or evaluation consultants. (One has a blanket item for consultative service that includes the area of evaluation along with other aspects of the program.) The smallest of these budget items is for five days of consultative service, while the largest provides $10,000 for continuing evaluative services during the first year of the project. In only one proposal is there specific provision for "evaluation materials and supplies." It may be that where no budgetary provision is made for time and materials for evaluation, these costs are included in general items for salaries and materials; however, it seems more likely that lack of budgetary provision for evaluation reflects a lack of focus on the evaluative process. The proposals that did include budget items for evaluation tended to be more definite in their plans for evaluating the project than those that did not.

What concept of evaluation seems to predominate in the proposals? Do the plans suggest that evaluation is a one-time procedure to measure the results of the project? Or do they indicate that evaluative data should be collected at frequent intervals and analyzed so that the evidence can be fed back into project development? Since the OE guidelines for Title III proposals stress the traditional concept of evaluation derived from experimental research design, it is not surprising that in most of the projects evaluation seems to be conceived as consisting entirely of assessment of end results. Of the 12 proposals that include attention to evaluation, eight limit their plans to this pattern. The other four proposals call for evaluation of final results, but in addition they provide for periodic evaluative sessions of participants, with feedback as a basis for shaping the next phases of project development. Participating teachers in one operational program meet each week and in another each month to discuss materials and teaching-learning activities they have employed, and to plan ahead. Three evaluative sessions are to be held during the project year by participants in a third operational program. The fourth proposal that provides for interim evaluation sessions is concerned with planning and pilot activities on a regional scale. The central group in this project meets each six weeks to review work done at the local level and determine any needed modifications in direction.

Recommendations

Do the Title III projects dealing with social studies share the weakness in planning for evaluation that characterizes Title III projects generally? Yes, if this small sample is indicative. The following conclusions apply to a majority of the proposals examined, and on most points the majority is a large one. That there are promising exceptions to each conclusion has been noted in the preceding section.

1. The provisions for evaluation are described so briefly and with such generality as to suggest that relatively little specific attention has been given to planning for assessment of the programs.

2. Such planning seems to be thought of as something to be left to outside consultants, rather than as an integral part of program development and a process that should be carried on cooperatively by project participants.

3. Clear relationships between particular project objectives and the procedures and instruments to assess progress toward each objective are usually not established.

4. There is little indication that the current literature dealing with evaluation in the social studies has been examined as a basis for planning for project assessment. Granting that there are many inadequacies in available strategies for evaluating social studies programs,

and that new approaches and instruments must indeed be created, it is nevertheless wasteful to ignore what has been done that could be drawn upon.

5. There is no specific budget provision for assessment activities in half of the proposals that included some statement about evaluation plans.

6. Evaluation seems to be conceived as measuring final outcomes, rather than as a continuing process that can feed back into program development. It is probably improper to cite this as an evidence of weakness in the projects' provisions for evaluation, however, since this is the conception on which OE guidelines for Title III projects were based.

There are some potentially positive features in the evaluation plans of the sample of proposals studied here.

1. Although achievement tests are the most frequently mentioned device for assessing pupil performance, other less traditional methods are also to be used in some of the projects. The approaches mentioned are not new, but they are used too infrequently. Perhaps Title III projects will encourage a wider, more effective application of them.

2. Several of the proposals included in the sample will broaden the base of project evaluation by obtaining evaluative reactions from various groups—teachers, pupils, and parents, as well as administrative

personnel—and will use a variety of instruments to do so. Efforts to achieve a broader base for judging the success of school programs have been made before, but perhaps Title III projects will encourage more schools to proceed in this direction and will develop more effective approaches for collecting the evaluative evidence.

3. Most of the proposals indicate that the development of new instruments for collecting varied types of evaluative evidence will be undertaken. If this is done and the new instruments are made generally available, social studies education will be enriched.

How might the provisions for evaluation in Title III projects in social studies be strengthened? Since standards for approval of projects are set by the OE guidelines, plans for project evaluation will reflect the emphases on evaluation that are developed in the guidelines. Greater attention to evaluation in the guidelines for the second year of Title III operations probably accounted for the slight improvement in this aspect of second-year proposals that was noted at the beginning of this paper. Further revisions of the section on evaluation in OE guidelines for future Title III proposals should be considered with a view to strengthening plans for project assessment. Such revisions might profitably focus on the following points.

1. THE CONCEPT OF EVALUATION AS A CONTINUING PROCESS THROUGHOUT A PROJECT, TO PROVIDE FREQUENT FEEDBACK AS A BASIS FOR DECISIONS

ABOUT IMPLEMENTING THE PROJECT, SHOULD REPLACE THE TWO-DIMENSIONAL MODEL THAT IS LIMITED TO EVALUATING FINAL PRODUCTS IN RELATION TO STATED OBJECTIVES.

A model such as the CIPP, which has been developed by Stufflebeam and his associates, might be adopted.[*] This model provides for specific and frequent evaluations of instructional materials and procedures in relation to project objectives, with feedback to provide a basis for decision-making, as well as for evaluation of final products.

2. THE GUIDELINES SHOULD MAKE CLEAR THE NEED FOR PROGRAM PLANNERS TO BECOME ACQUAINTED WITH EXISTING APPROACHES AND INSTRUMENTS FOR EVALUATION OF PROGRAMS IN THE FIELD WITH WHICH THE PROJECT IS CONCERNED, TO THE END THAT AVAILABLE RESOURCES MAY BE DRAWN UPON IN FORMULATING IMPROVED APPROACHES AND INSTRUMENTS THAT WILL BE APPROPRIATE FOR THE PROJECT.
3. THE STATEMENT IN THE 1966 GUIDELINES CONCERNING BUDGET PROVISION FOR EVALUATION ACTIVITIES SHOULD BE STRENGTHENED TO ENCOURAGE PROPOSAL FRAMERS TO MAKE SPECIFIC AND ADEQUATE PROVISION OF RESOURCES FOR PROJECT EVALUATION.

---

[*] Daniel L. Stufflebeam, "The Use and Abuse of Evaluation in Title III", 1967 (mimeographed).

URBAN, METROPOLITAN AND RURAL EDUCATIONAL DEVELOPMENT[*]

The sections dealing with Evaluation that appear in the proposals for Supplementary Educational Centers are generally rather perfunctory, and two reasons probably account for this situation. In the first place, the Center's programs tend to be broad and rather vaguely defined. They usually propose to create new courses of instruction with new teaching materials, or to train teachers and counsellors for new roles. Thus they do not lend themselves to an experimental design, with experimental and control groups of students and statistical tests of various hypotheses.

In the second place, the concept of evaluation used by the people who have planned the projects is generally limited to the evaluation of the outcome of the project. This is good, but not enough.

Evaluation procedures can be used to assist in making decisions at four stages of a project.

1. Before the project is begun, to decide what needs should be met by the project.

2. At the beginning of the project, when its detailed procedures are being planned, and its personnel selected.

3. While the project is in process, in order to improve performance and correct mistakes.

---

[*] Prepared by Robert J. Havighurst, John Mosler Chair in Urban Education, School of Education, Fordham University.

4. At the end of the project, when its results can be evaluated and decisions made about next steps.

In this report the writer will attempt to show how evaluation procedures might be used at all four stages of two hypothetical projects, one of them (A) being a supplementary center in an urban-metropolitan situation which is expected to improve occupational education and guidance in a metropolitan area, while the other (B), is an educational facility to serve a multi-county rural area of about 100,000 population.

A. This is a metropolitan area with a population of 500,000. It contains three counties, has a varied racial population, a growing industry and commerce. There are about 60 local school districts, and three county school units.

B. This is a six to eight county area, with about 60 local school districts. Population is stationary or declining in all except one county, which is growing due to the recent arrival of several corporations which have located branch factories near the county seat. The principal industries of the past, agriculture and forestry, are now fairly profitable but the small farms of the past have been combined into large, mechanized farms requiring capital to operate. Therefore families have been moving into the towns and out of the area to larger cities for the past 20 years.

Evaluation Procedure for Stage 1. A Study of the Setting, or the Context of the Project

	A	B
Information to be Supplied	Socioeconomic data Occupational Distribution Achievement levels of school children Attendance in school—public and private—in relation to age and sex Results of surveys that have been made	SAME
Decisions to be Made	What kinds of base-line data are needed for a useful record of the project? What are the principal educational needs of the community?	SAME

Stage 2. Getting the Project Started

	A	B
Information to be Supplied	How well prepared is the agency to carry on this project?	SAME
	Survey of available personnel	
	What facilities are needed? Are they available?	What consultant help is being used?
	What consultants will be used?	How is the project related to the State Department of Education?
	Plans for cooperation with other agencies. (School districts and other organizations)	

Decisions to be Made	Choice of Director Selection of Staff Location of Vocational Schools Will the new faculty be specially trained for the new program?	Choice of enrichment programs for the area. How made? How will new personnel be selected? Selection of a program design. Location of the various parts of the program.

Stage 3. Carrying On the Project

Information to be Supplied	Provision for feedback to Director and staff: from local industry; from students. Data on dropouts from the new courses. Provision for recording unexpected events relevant to the project.	Feedback from County Committees Data on changed holding power of schools Data on attitudes of youth Reports from Faculty
Decisions to be Made	Correction of errors Dropping of unsuccessful courses.	Same

Stage 4. Evaluating the Project. Deciding on Next Steps

Information to be Supplied	Measurement of output. Number of students trained. Jobs obtained by them. Attitudes of employers. Interpretation of outcomes	Number of students served by project Changes in proportions graduating from high school Attitudes of students and their parents

Decisions to be Made	Next steps. Shall the program be continued? Use of an Evaluation Committee. What modifications are to be made? Shall the program be extended to other schools or areas?	SAME

The proposals for Supplementary Educational Centers which the writer has read are generally fairly adequate with respect to Stage 1. They present the setting with a description of the community or the area to be served and the students to be served. Sometimes they report a survey or study of educational needs.

It is at Stages 2 and 3 that the proposals are most likely to be inadequate. Seldom do they show evidence of careful evaluation of the capability of the school district or other agency to perform in the area of the project. They do not often comment critically on the personnel who will carry on the project. They do not show how they will proceed systematically to find the new persons needed, and to give them adequate preparation through work with consultants, visits to other communities, etc.

There is almost nothing to show how the project will be evaluated while it is in progress, so that it can be improved, errors corrected, and progress reports made. This is especially important in innovative work, where one must expect to learn from experience as the project progresses. A periodic stock-taking is desirable, using local staff and community committees as well as outside consultants.

At the final stage the proposals are somewhat more clear on the procedures for evaluating the results. However, they seldom show evidence of thought about the next steps after the project funds have been used up. What prospects are there for continuing and expanding the program it if proves to be successful? Will there be a responsible committee or commission to make recommendations to the school board or other policy-making bodies?

The concept of evaluation dealt with here requires an evaluative effort at every stage of the project, and it cannot easily be described in a brief section toward the close of the proposal. Rather, as the description of the project proceeds, there should be a description of plans for collecting the information and making the kinds of decisions suggested in the outline given above.

ORGANIZATIONAL FLEXIBILITY[*]

When one studies the evaluation sections of many of the Title III proposals, a question immediately is raised: What information will exist at the end of the pilot exemplary or innovative program that will allow for an immediate decision to make the innovation pandemic or universal? This means not only for the rest of the school system involved in the project, but for expansion to other school systems throughout the United States which are in a never-ending search for programs that will more sufficiently be up to the task of the education of our young.

Unfortunately the studies of and visits to various Title III projects dealing with innovation and change in the areas of team teaching, independent study, the individualization of instruction, various approaches to collaborative teaching, conceptualizations concerning nongraded programs from the pre-school on through the twelfth year, and other variations of these things, do not lead to the acquisition of knowledge, or to that collection of research findings upon which one can base a more intensive action. Rather than seeking implementations and collected evidence, for the most part evaluations are still using such techniques represented by the following statements taken from the proposals:

"A careful record of each project activity will be maintained. A careful record of student and adult participation will be maintained and summarized. Pre-test and post-test will be used to determine exact results in remedial reading classes. Statements from teachers stating the value of individual situations will be gathered. Statements from teachers concerning the advantages and disadvantages will be gathered. Persons viewing demonstrations will complete questionnaires comparing what they saw with traditional type classroom activities. Recorded comparisons shall take place. A survey technique will be maintained where we have a student-parent diary. Children will be tested at the beginning of school with an achievement and intelligence test and again at the end with an achievement test. Teacher aides, teachers, supervisory personnel, college educators will record observations during the year and through written reports and oral discussion evaluate the project." And so it goes.

---

[*] Prepared by Maurie Hillson, professor of education, Rutgers—The State University of New Jersey.

Very little of this material would stand the scrutiny of any scientific investigation. More importantly, very little of this material would be readily useable for making the educational decisions that would be necessary for successful expansion or implementation. If educational evaluation is moving toward the position of becoming a process during which data is collected, decisions made to implement, and modifications of programs are suggested, then the situation as it presently stands in all too many of the operable Title III programs is going to be difficult to deal with in this context. There are several reasons for this.

The adequacy of any evaluation section of any particular proposal is going to be in direct relationship to the precision and insight with which the objectives or the prescribed proposed models of the innovated exemplary activity are stated. As one studies the various proposals, it becomes a critically painful fact that very few people who are presently being funded and working on the proposals have a clearly defined statement of the emphases concerning the various aspects of teaming, collaboration, nongrading, and other vehicles. The procedures and approaches to evaluation as they presently stand are merely technical aspects of measurement. These are not wholly appropriate to the kinds of innovative and exemplary programs and proposals that are either subsumed in the titles of the proposals or discussed in broad terms in the narratives of the proposals.

One is led to the interpretation that the more difficult area of analysis for the awarding of the proposals should be not so much in the evaluation schema, but rather in relation to the specificity of the stated objectives and the written defense of intellectual as well as implementation activities concerning the kinds of things that are going to be attempted in the exemplary or innovative program. Unless one has a very clear picture of the various kinds of possible models that break with the customary way of doing things, and unless one has a clear picture of the various processes and vehicles that will bring about some particular change, then what may result is a proposal based on a set of mere verbal gymnastics recasting only what is a very traditional program.

The adequacy of the various approaches to evaluation, the procedures involved, the techniques being employed, do not really appear at this particular stage of the Title III activity to indicate that they are strong to the point of underpinning the innovative impact in American education called for by Title III.

There are several kinds of things that one must think about in strengthening the various proposals. One cannot simply strengthen the evaluation section of the proposal without giving basic consideration to the particular objectives that the proposal has set for itself in terms of the innovative and exemplary structure. Future proposals should carry specified evaluative techniques for probing the extent to which specified objectives of the proposal have been attained operationally. For example, if one contends that a basic vehicle for creating a new process of education is "teacher team planning" because it affords the opportunity to the teachers to make relevant decisions about learning, the opportunity for the construction and reconstruction and modification of curriculum, as well as many other things concerning the way the child will be taught and how he learns, then the evaluation section must carry a particularized kind of activity that evaluates the outcomes of "teacher team planning" as compared to teaching in a self-contained situation where no teacher team planning is available. There should be consideration of the evaluation activity for the opportunity to change the objectives. For example, in view of the fact that "teacher team planning" may bring with it certain unforeseen consequences, or ancillary activities, which at the moment of planning were unknown to the group suggesting that innovative type of activity, evaluative procedures to measure these new items must be established. I think that it is very important that each item of evaluation be stated in terms of the resultant behavior expected. The proper instruments must be created and employed so some realistic measure can be taken. I think the mere suggestion of pre- and post-testing and standardized tests although necessary and helpful are often limitations to, rather than activities that offer insight for, educational change. There is no question that certain aspects of growth need to be measured by various kinds of tests and inventories. But, more importantly, the contexual operation of a program needs to be evaluated so it will provide immediate information that will allow for the continuing change and implementation of the major emphases that are deemed valuable and that should mark the educational programs of our country.

Evaluation must be thought of as a program that serves all functions of a proposed activity in an effective and precise manner. To be effective and precise, evaluation must serve several functions. The functions of determining whether or not various factors are the reasons for various deficiencies, whether or not the functions are providing reliable collections of evidence about the students concerning progress, whether or not the functions indicate the extent to which various students in the program are making progress toward the stated goals, and whether or not the functions indicate the successful manners by which various kinds of activities should take place, vis-á-vis grouping, individualized instruction, skill impact stations for teaching basic skills, and many other concerns relative to the learning-teaching act and educational setting.

One cannot be sure that an effective program of evaluation has been established unless criteria are established that indicate rather specifically what the adequate aspects of the evaluation program are. It seems relatively important to state that if the Title III programs are to be valuable in terms of their effectiveness, then the evaluation section must be basically involved in at least establishing the following guidelines and activities in the proposals:

(1) The evaluation section and activity should be quite comprehensive. It is very important that desired goals not be omitted just because there may be difficulty experienced in appraising these goals objectively. A comprehensive evaluation insists that all of the major objectives be stated.

(2) The evaluation section should be so set up that the

educational objectives are stated in a realistic way. The evidence that is collected should indicate the extent to which the changes in student behavior have taken place or are achieved. This is extremely important.

(3) It is assumed that the statements of objectives and the analysis of those objectives will represent more meaningful resultant evaluation if all teachers participate, or all persons who are involved in the program, participate in establishing, analyzing, and developing various ways that the evaluations will be made.

(4) It follows that the people involved also would be in the continuous process of collecting data that would allow for a continuous as well as a contiguous evaluation of the growth that is taking place in the particular program of change.

After a careful analysis of various Title III projects, in all candor it may be said that the most important thing in evaluation concerning these projects should be the concept of reality. Some difficulties and problems of evaluation now may be due to the fact that the statements that are made concerning aims and objectives are somewhat grandiose. It is impossible for this writer to accept the notion that rural and rather conservative school districts are overnight, because of some federal money, going to move from their conservative status to becoming the lighthouse districts of the American public education system. It becomes very important that the evaluation program be conceived in terms of the local situation out of which the proposal arises. The expectancies, the norms, the instruments to be developed, the content of the various kinds of tests to be used to measure the outcomes, the questionnaires that will be used to evaluate or probe the depth or extent of participation and the resultant changed behavior, are all basically relative and related to the situation out of which they arise. This becomes an extremely important point. It allows for differential application of support throughout the country. This would be very important in improving the programs, proposals, stated objectives, and the evaluation sections of the present Title III projects.

Finally, there should be some kind of creativity in the area of evaluation. The basic utilization of many techniques to find out what has happened is essential if one is to obtain various kinds of data that will indicate changes that have taken place. For this reason I think that evaluation models that have an open-ended quality to allow for the measurement of changed directions of activities because of ancillary or unforeseen consequences arising during the program, are far more fruitful than precise statistical research designs that may be quite sterile and bereft of flexibility. These are aspects that could prove helpful in strengthening the present Title III proposals, those under consideration, and those that are operable.

Recommendations

Several things need to be considered, but basically the following represent at least minimal requirements for assuring stronger, more adequate, and more relevant evaluation procedures:

1. THE EVALUATION SECTION OF THE PROPOSAL SHOULD BE TIED SPECIFICALLY TO THE STATED OBJECTIVES IN THE PROPOSAL. FOR EVERY STATED OBJECTIVE THERE SHOULD BE A SPECIFIED EVALUATIVE ACTIVITY THAT WILL DETERMINE THE EXTENT TO WHICH THE OBJECTIVE HAS BEEN ACHIEVED OR IMPLEMENTED.
2. OBJECTIVES SHOULD BE STATED IN TERMS OF CHANGED BEHAVIOR OR IMPLEMENTED EDUCATIONAL VEHICLES. THE ANALYSIS AND MEASUREMENT OF THESE OBJECTIVES SHOULD ALSO BE DETERMINED BY DESCRIPTIVE ACTIVITIES THAT SHOW THE EXTENT THAT BOTH THIS BEHAVIOR AND THESE VEHICLES HAVE BEEN ACHIEVED.
3. THE OBJECTIVES SHOULD BE SO STATED AND SO MEASURED THAT WHERE UNFORESEEN CONSEQUENCES DEVELOP BECAUSE OF WORTH-WHILE OPERATIONAL PROCEDURES, OBJECTIVES CAN BE ALTERED AND THE

PARTICULAR FINDINGS USED TO INDICATE OTHER DIRECTIONS OR MODIFICATIONS OF ACTIVITIES PRESENTLY OPERABLE.

The imagination and insight brought to bear in this area of concern can lead to the collection of relevant and meaningful materials that can readily serve others in moving forward to advance creativity in education.

PUPIL PERSONNEL SERVICES[*/]

Projects funded since January 1, 1967, differ from earlier ones in their approach to evaluation and in their procedures. The earlier evaluations tended toward generalities, descriptive data, and a small variety of instruments, frequently limited to achievement tests. The later projects may be described, although not universally, as:

1. Displaying the language of evaluation. The words that have appeared in evaluation literature and at evaluation conferences come out prominently in many proposals.

2. Evidencing an acceptance of the evaluation concept of measuring progress toward objectives. Pre- and post-measures appear. In some projects there are references to an initial data base.

3. Using a greater variety of instruments within the evaluation of a single project. One project has eight areas of instruments, and most projects have escaped from the limitation of descriptive data exclusively, such as the number of participants.

These generalized comments describe movement toward more relevant and meaningful evaluations. But it is only "movement toward;" the evaluations are far from the level that should be achieved. Only one project has reached the point in evaluation procedures of a feedback, change, and decision-making process. Two projects had a model for their evaluation.

[*/] Prepared by Arthur A. Hitchcock, professor of education, State University of New York at Albany.

Probably the greatest problem in the evaluations is the absence of a plan, with the exception of the two projects. In most instances the evaluation "plan" is a list of instruments to be administered or descriptive data to be collected. This flaw is extremely serious.

Although projects after January 1, 1967, show positive characteristics as stated above, yet the lack of evaluation plans almost inevitably carried the concomitant that some of the instruments proposed are exceedingly dubious. One wonders how they could contribute anything to the evaluation.

Specific Description

1. Program of Evaluation: Evaluation programs are woefully lacking. The 20 proposals reviewed range from no program for evaluation to two projects that have realistic, workable programs. If considered on a scale, ten of the projects would fall at and near the "no program" end of the scale; five would cluster toward the center of some semblance of a program; two have programs; three would not be included because they are planning grants.

Most evaluations are simply lists of things—instruments, data, conference plans—without their being drawn together into a schema. Three of the evaluations show that the procedures will be left in the hands of others such as an educational laboratory.

Nearly half of the evaluations make some reference to test usage or data collection pre and post. This is suggestive of the beginning of a plan. But in analyzing the proposals in detail, it is found that there are only infrequent comments to connect pre and post activities to specific objectives. In other words, the beginnings are present but there is a long distance to go.

Without plans for evaluation, these projects do not envisage feedback and change. With the exception of the two projects, they are at the point of showing varying degrees of understanding of the concept of behavioral objectives and of measurement toward them, but not in a planned package, and not to the level of the feedback and change concept. The control group concept also shows the lack of a plan. Control groups are mentioned only rarely and never as part of a definite plan.

Every project includes consultants, but the nature of them varies considerably. Seven of the projects expect to use state and federal educational personnel in the evaluation, but only one project shows any plan for their utilization.

2. Instruments for Evaluation: Descriptive Data. Virtually every project calls for descriptive data in the evaluation. Descriptive data are actuarial counts. The types are:

Number of participants
Amount of participation
Number of dropouts
Number of referrals (for counseling, discipline, medical aid, etc.)
Number of actions (such as counseling interviews, home visits, parent visits to school, production of materials)

Number in post-project action (such as entering college, taking higher level courses)
Case studies (both number and content)
Health records
Attendance records
Socio-economic data, usually about family
Teachers' marks

Objective Instruments. Standardized instruments sold by test companies are used in practically every evaluation. Objective instruments constructed locally are found in half of the evaluations. The types in the projects reviewed are:

Achievement tests
Mental Ability Tests, group and individual
Attitude Scales (for students, teachers, administrators, community)
Interest Inventories
Reading Diagnostic and Achievement Tests
Tests of Learning Disorder
Personality tests and scales, projective and paper-and-pencil
Motor Development Tests
Questionnaires
Opinionnaires
Socio-grams
Anecdotal Records of Behavior
Checklists of Behavior
Student Autobiographies and Diaries, channeled in a format.

Subjective Data. In addition, in five of the projects data are to be gathered that will come from interactions of people. The types mentioned are:

Conferences (of lay citizens or professionals)
Juries
Interviews
Observations

The variety of instruments has grown enormously from the earlier projects. Unfortunately, without plans for evaluation the efficacy and cogency of the instruments can only be surmised.

3. Other Aspects of Evaluation: The idea of behavioral objectives is catching on, but the idea of criteria related to objectives is not. Nor is there any evidence that the specific objectives of a proposal are intimately connected with the student needs from which the project grew. In only one project out of 20 was there assurance that the evaluation program started with student needs. In another project this connection might be assumed.

The projects reviewed show signs of improvement in evaluation. Hopefully the next ones will improve further.

Analysis and Interpretation

First, the evaluations are beset with psychological fears. A feeling emerges in studying the projects that a big, bold idea and plan of action—or even a little, timid one—suddenly dissipated itself when the evaluation section came up. I think that there are several reasons for this:

1. The truism, still true, that educators are unaccustomed to evaluation and the thought is scarey. A person in education is accustomed to action and interaction, and he can advance boldly on this front, and even advertise his action plans and procedures. But he feels uncomfortable with evaluation. This is something, he thinks, that requires more competence than he has.

2. A certain defensiveness exists that evaluation cannot do justice to the project, anyway. The educator thinks that evaluation at best will touch only a part of what he is accomplishing and, therefore, it is something to be feared.

3. A fear that the colossus in the state education agency or in the USOE will conclude from the evaluation that the project is not as good as it really is. This further inhibits a project developer from performing as well on the evaluation as perhaps he could.

It is important to understand psychological fears of evaluation in order to understand what is not going right.

Second, the evaluations are not creative. In fact, most of them are incredibly dull. If the proposals for action were as unimaginative as the evaluations, they would not have been approved. This may well be a product of the failure of evaluation plans, as noted in the description above, for the process of working out a plan can stimulate thinking in more creative ways.

Third, the evaluations show a lack of understanding of evaluation as a process. With only one project definitely showing a feedback process, and another implying it, one is forced to the conclusion that most leaders in the projects reviewed have not grasped this concept. They are at the stage of seeing the connection between objectives and measurement, but as a static thing, not as part of a larger process.

Fourth, evaluations are encapsulated by experimental design.

Stufflebeam's admonition in his paper on the CIPP design is borne out particularly by one project in which university researchers drew up a fine experimental design, but missed the point of evaluation process. It must be said that most of the evaluations show signs of experimental design. This is understandable because this is the kind of preparation that project leaders have had.

Fifth, the evaluations are not reaching the heart of the matter. In projects broadly in the pupil personnel area, changes in behavior—stated with specificity—will inevitably be the objectives. Yet the evaluations, for the most part, are devoid of any incisive move into the very heart of what the projects are about. Without good plans, the evaluations skirt around the outside, and never reach into the heart of the matter.

Sixth, the better evaluations had outside consultative help. This point has implications both for the quality of some of the assistance and for the needs of project leaders in developing evaluations. In fact it states so much about the evaluation problem at the local level that one must become terribly concerned about speed of progress in developing better evaluation plans and practices, about preparing people to do evaluations, and about the persons who should prepare and even carry through evaluation plans.

Seventh, the format for narrating the evaluation in the proposal tends to discourage project leaders from viewing evaluation as a process and as a part of the whole project. There should be a format that would encourage thinking about evaluation plans in connection with student needs, environmental conditions, and specific objectives of the project. At present the evaluation section is one of the separate parts of the narrative in the application, and this encourages thinking about evaluation as a separate part.

Eighth, the evaluations show improvement over earlier ones. In other words, learning has been occurring. This should suggest avenues for the future. But the evaluations that were studied are such a distance from the level they should be that more drastic solutions must be found than have appeared thus far. The project leaders indicate that they want to do well, that the failures to do so are more products of the conditions stated in this analysis than in the desires of the leaders. The future of evaluation should be built upon this positive condition.

Ninth, the evaluations in these projects illustrate the trouble in the world of evaluation. There has been progress. True. But these projects show lack of cohesiveness of evaluation with the project, lack of comprehension of what evaluation is, lack of tackling the toughness of evaluation; namely, facing up to exactly what we are doing.

Finally, models could be applied to several of these evaluations. One project is close enough that the CIPP model (Context, Input, Process, Product Evaluation) could be used without difficulty. Two other projects could follow the CIPP model with some alterations that probably could be made. All of the other evaluations would have to be reconstructed in varying degrees. (The attempt will be made actually to follow the CIPP model in the one project by working with the project leader.) However, it is believed that other models may also be applicable. The point of first importance is bringing project leaders to an understanding of the evaluation process. Second is the action of creating such a process within the project, probably by the use of models that can be modified for particular conditions. Surely various models will emerge, and it appears now that they will be based upon the process concept of evaluation.

Recommendations

This paper is focused on the evaluation of projects broadly classified as pupil personnel services. Yet, it is impossible to view evaluation apart from projects themselves, for evaluation, like the total project, contributes to or inhibits innovation; serves or disserves national strategies and concerns; enhances or reduces the thrill of movement of human lives. Evaluation is part of the Title III complex. These recommendations are directed at evaluation but they are constructed with the totality of projects in mind.

1. DEVELOP MORE VALID, RESPONSIVE, AND CREATIVE EVALUATIONS OF PUPIL PERSONNEL SERVICES AT THE LOCAL LEVEL BY:
   a. Expanding the training of project leaders.
   b. Providing competent, specialized consultants at the local level during the planning stage.
   c. Creating a very small group of highly competent pupil personnel persons—two or three—to maintain a national view and grasp of evaluation activities, based on knowledge of local programs. (This group would be used also in advising the USOE on Title III projects.)

It is obvious from this study of evaluations and from knowledge of working projects, that project leaders want to perform better on evaluation. But they need further training, the kind that helps them with their own specific live projects. And, they need the kind of help that can come only from individuals working with them. The national view recommended is important, also, to move evaluation plans and procedures along as rapidly as possible, but now most project leaders have a great deal to learn before they can move to the level they desire.

2. DEVELOP LARGE-SCALE NATIONAL, LONG RANGE EVALUATION PROGRAMS IN PUPIL PERSONNEL SERVICES.

Such a program would necessitate fitting local evaluations into the national schema. This could be accomplished without harming local initiative and variation. This will make it possible to pump new evaluation concepts and procedures into local projects for adaptation, thereby encouraging self-renewal. It will also contribute to better knowledge about what is happening in pupil personnel services. Such a program should be set initially for perhaps five years.

3. CONDUCT A VERY HIGH CALIBER, INNOVATIVE STUDY OF NEEDS IN THE EVALUATION OF PUPIL PERSONNEL SERVICES, AND HOW TO MEET THEM.

The pupil personnel area is especially difficult for evaluation. It is concerned with the totality of individuals and much of the work is in the affective domain. The effects of processes and programs on behavior frequently are difficult to ascertain in short time periods. Yet, these very conditions are exciting and challenging. The study would be concerned with the adaptation of models and perhaps guidelines for new models, with kinds of new instruments that are needed for evaluation, with strategies for testing procedures under varying conditions. The evaluation of pupil personnel services is not only one of the most difficult, it is also in bad condition. It needs the push that could come from such a study.

This study should be done by a very small group, and with a limited time span. But it must be a hard-hitting study, for the needs are urgent.

ROLE OF THE STATES[*/]

The CIPP model provides a suggestive approach for the discussion of the role of the States in PACE evaluation, calling attention to the fact that evaluation is a continuous process and is an integral part of the States' responsibilities in the administration of Title III.

1. Evaluation of Statewide Needs

The State's evaluative role starts with its assessment of statewide needs and with its establishment of priorities for meeting those needs. Most departments make such an assessment and set priorities for the overall programs of the department. It may be desirable, in addition, to specify which needs are to be the particular focus of Title III projects. In determining PACE priorities, account should be taken of reports from ongoing Title III projects. In this way, information generated by other stages of evaluation will be fed back to assist in formulating and reformulating basic priorities.

2. Evaluation of Preliminary Project Ideas

The next stage in evaluation comes when ideas for possible projects are first broached to State Title III staff members either in formal Letters of Intent or in informal discussions. The first test to be applied is whether the idea seems to have sufficient intrinsic merit to be worth developing into a full-scale proposal. This is a judgment which the staff member should be able to make on the basis

---

[*/] Prepared by Norman D. Kurland, director, The New York State Education Department.

of his own knowledge. If he has any doubts, he may certainly consult with his colleagues or specialists in the field, but he should be able to make this judgment on his own. Dependence on staff judgment of this kind at this point is one of the reasons why Title III requires highly qualified staff members.

If the idea appears to have intrinsic merit, then certain extrinsic criteria should be applied, such as whether the idea is already being developed in another project; whether the personnel would be available to carry out the idea if funded; whether the district or its region already have so many Title III projects as to reduce the probability of further grants except for exceptionally outstanding proposals; and, finally, how the idea relates to the priorities established for Title III. Information on these matters should be presented frankly to the potential applicant so that he may judge whether it is worthwhile to develop a full proposal. If he decides to proceed, he can be greatly helped if he and the state department staff member can work out an approximate budget to be used as a guide in developing the project budget.

3. Proposal Evaluation

The criteria for review of project proposals are given in Appendix C of the PACE Manual. State evaluators should now take into account the recommendations on evaluation made in the other papers in this study.

4. Program Monitoring (process evaluation)

There are at least four occasions for the monitoring of ongoing Title III projects:

a. Whenever a Title III staff member or other Department member has occasion to visit a district in which there is a Title III project he can get an informal reading on the project. Special problems or outstanding accomplishments are particularly likely to come to his attention even without any intensive review of the project.

b. Acceptance of invitations to visit projects on special occasions or for monitoring purposes provides another opportunity to find out how projects are going.

c. In response to requests for amendments to contracts, information on the project will usually be submitted and may be requested.

d. Development of the State's recommendation on continuation of the project at the termination of each grant period will require analysis of progress reports and, generally, a site visit.

The primary questions to be answered whenever any of the above contacts with the project are made are the following:

a. Is the project being carried out as indicated in the initial proposal and contract? If not, and if deviations have not been approved, are they serious enough to warrant attention by the Title III office?

b. Is the project on schedule?

c. Are the funds being handled properly? (books in order, bills paid on time, etc.)

d. Is the project being received well in the area? If not, what are the sources of dissatisfaction and is the project staff aware of them?

It should be stressed that process evaluation is not the occasion for determing the worth of the project. This question was first decided when the proposal was approved and will be a matter to be considered again when the objectives of the project have been achieved. The basic purpose of process evaluation is to help strengthen the project and increase the probability that the objectives will be realized. The only basis for terminating a project while in process should be the judgment that it is so far from making adequate progress toward its objectives that no amount of reasonable adjustment will make possible achievement of the objectives.

5. Product Evaluation

This is the phase of evaluation when the outcome of the project is judged in relationship to the original objectives. Data from the project's own evaluation should be the primary source for judgment at this point. If the evaluation design was sound, no further data should be needed. However, in many cases the State staff may wish to supplement the evaluation reports with on-site visits by Title III staff, staff from other department units, or outside specialists.

The results of this stage of evaluation should be used by the department to determine what recommendation to make on the idea; whether the idea is sound and feasible enough to be recommended for adoption elsewhere; whether, though not yet ready for wide-scale adoption, it still is promising enough to be worth additional development; or whether the idea did not fulfill its promise and should not be encouraged further. Whatever the decision, it should be formally reported to other Department units, to the field and to the U. S. Office of Education. Where adoption or further development is indicated, the Title III Office should play an active role in disseminating the idea or encouraging its further development. In part, this will be done by using the results in the reformulation of the priorities for Title III. Thus the evaluation loop is closed. Information from complete projects provided guidance in determining the lines along which new projects should be encouraged. New projects can build on previous experience or be developed to deal with new problems.

SCHOOL-COMMUNITY RELATIONS[*/]

Local educational agencies in preparing Title III proposals are challenged in the PACE Manual to develop imaginative solutions to educational problems; to utilize more effectively research findings; and to create, design, and make intelligent use of supplementary centers and services while translating the latest knowledge about teaching and learning into practice. In meeting this challenge, planners are confronted with the need to develop new school-community relationships by incorporating into their programs potential assets of the community which have not traditionally been vitally involved in the formal educational process. Further, they are called upon to explain the plan for evaluating the effectiveness of the program and the extent to which the objectives have been accomplished.

The sample of proposals chosen for this review are concerned mainly with school-community relations, and this review focuses upon the evaluation section of the proposals—the approaches, techniques, and general design for evaluating the effectiveness of the planned activities. Particular attention was given to determining:

1. Degree of clarity and comprehensiveness of the over-all evaluative design.
2. Relationship of the evaluative process to the operational process.

[*/] Prepared by John W. Letson, superintendent of schools, Atlanta Public Schools.

3. Innovative aspects of the design.
4. Extent of concern for the feasibility of the program.
5. Relationship between budgetary commitments and design requirements.
6. Relationship between focus of the evaluation and qualitative or quantitative changes.
7. Compatibility of evaluative procedures with stated objectives.
8. Identification of units of change and variables involved in the change process.
9. Appropriateness of techniques for measuring achievement of objectives.

Uniformity in analyzing the various proposals was gained by applying an instrument developed by the reviewer, consisting of items pertinent to the nine general areas identified above. The degree to which the evaluative plan supplied information concerning these areas was recorded on a five-point rating scale for each item of the instrument. The scale was designed so that a rating of "1" indicated that the proposal did not contain information concerning the particular item, the information was not sufficiently clear, or it was too broad, sketchy, or general. A rating of "5" indicated that the information was very adequately stated and showed that the planners had given sufficient guidance for future action on the point in question. Care was exercised in making these value judgments because of the recognition that experimental design conditions and inflexible operational procedures are not appropriate evaluative features for many Title III programs, especially those focusing on school-community relations. In making the ratings, emphasis was given to the procedures and techniques which will be used to determine the measurable or observable effects and to the appropriateness of the procedures for the stated objectives.

In addition to analyzing the internal features of the evaluative section, comparisons were made of design differences of the evaluative procedures because of the increased emphasis on evaluation evidenced in the successive editions of PACE and because of the content of the programs.

Proposals reviewed in the sample included those funded in each of the successive grant periods, extending from November 10, 1965, through January 15, 1967. Emphasis on evaluation was not evident in PACE until mid-1966; therefore, a comparison was made between proposals funded before January 15, 1967, and those funded subsequently. A comparison concerning the content was made between proposals dealing specifically with utilization of community resources or the development of school-community relations and those dealing with in-school activities, such as the teaching of reading.

Description

The sample reviewed consisted of 19 proposals selected at random from those submitted to the reviewer. Of these 19 proposals, 3 were for planning grants; 11, for operational grants concerning school-community relations; and 5, for operational grants concerning in-school activities. The number of operational proposals in the sample which were funded before 1967 was about equal to the number funded in 1967.

NUMBER OF PROPOSALS IN THE SAMPLE ACCORDING TO DATE OF FUNDING, TYPE OF GRANT, AND CONTENT

		Date of Funding		
Type of Grant	Content	Before 1967	In 1967	Total
Planning	School-Community Relations	1	2	3
Operational	School-Community Relations	5	6	11
	In-School Activities	2	3	5
	Total	8	11	19

Planning Grants: (School-Community Relations)

Three of the 19 proposals selected at random from those submitted to the reviewer were for planning grants in the area of school-community relations, which will finance co-operative efforts to provide improved educational and cultural experiences for pupils, teachers, parents, and others. A brief description of each of these three proposals follows:

1. Mid-Monmouth Educational Council (Long Branch, N. J.) The council will serve to expand educational and cultural opportunities for students, to help teachers to improve instruction, to conduct significant research projects, and to explore creative educational

innovations at the high school level. Participating council members consist of four local educational agencies which will plan for collective utilization of their resources and for a demonstration program designed to serve as a model for future co-operative efforts.

2. Community-School Study Centers (Webster Springs, W. Va.) The primary objective of this project is to determine the most efficient and functional way of making available library resources and cultural experiences to elementary and secondary students during and after school hours through community-school study centers in an isolated rural area in which no public libraries and/or cultural facilities are available.

3. School-Community Educational Service Center (Lucia Mar Unified School District, Pismo Beach, Cal.) The purpose of this project is to explore the possibilities of establishing a school-community educational resource center which will provide (1) an organizational structure for co-operative solutions of peoples' problems; (2) educational materials and equipment; and (3) a staff which will render innovative, coordinated educational services to children, parents, teachers, prospective teachers, welfare and law enforcement agencies, family service agencies, mental and public health agencies, vocational rehabilitation agencies, and professional and fraternal organizations.

Operational Grants: (School-Community Relations)

Eleven of the 19 proposals selected at random from those submitted to the reviewer were for operational grants devised to utilize co-operatively, innovatively, and creatively all available resources of schools, communities, metropolitan areas, and even multi-county school districts to provide improved educational, personal, and cultural opportunities for students, teachers, parents, and others. A brief description of each of these eleven proposals follows.

1. A Co-operative Project to Provide Supplemental Services to a Group of Elementary and Secondary Schools of New Mexico (Santa Fe, N. M.) The purposes of this project are (1) to provide to elementary and secondary schools certain essential services now entirely lacking or inadequately provided, (2) to provide school opportunities for the professional improvement of the teachers of the area through inservice programs, (3) to provide eventually quality education and broad educational opportunities for the children of the area, and (4) to provide a central facility or clearing house for a group of 26 schools which will enable them to engage in a variety of co-operative activities. Some of the activities will be in curriculum development, psychological services, music and cultural programs, special education, planetarium programs, adult education, and data processing.

2. Alameda County Supplemental Services Center for

Assurance of Community Creativity (Hayward, Cal.) The center is designed to identify and assess critical educational needs within the County; to identify and utilize regional and national resources in conceiving, planning for, and implementing exemplary solutions to problems; and to evaluate potential and ongoing Title III programs within the County. The center, through its research and development activities, will provide program intergration and support to urban and suburban agencies for effective utilization of regional and national resources and for efficient implementation in the various school districts. It is envisioned that fiscal support of the center will be expanded to include philanthropic agencies.

3. A Co-operative Project Among Teachers, Schools, and Industry for Continued Development of Means to Improve Learning (Oak Park, Ill.) This project is designed to continue a pilot program made possible from the Knapp Foundation and a former Title III grant in order to develop a library-located instructional resource center capable of storing electronically vast amounts of information and making that information instantly retrievable for individual or small-group instruction. Ampex Corporation is providing systems engineers and a research staff to modify and to develop the electronic equipment.

4. Community School Program (Berlin, N. J.) The project is designed to involve the total community and to utilize the school

plant for recreation, physical education, and academic enrichment. The program will be for children and adults and will be conducted afternoons, evenings, Saturdays, and during the summer. The major objective is to have the activities and interests of the community centered in the school.

5. Library and Instructional Resources and Technical Services Center (Media, N. Y.) The project is planned to establish a regional library support operation in an eight and one-half county area. It will be innovative in that it will be the first library support center in New York State in which a Public Library System will contract for services from a School Library Support Center.

6. An Exemplary Comprehensive Family Life Education Project Including Training for Leadership (Amherst Central High School District, Buffalo, [Snyder] N. Y.) The major emphasis of this project is to employ a positive group-education approach to help meet growing needs for Family Life Education in Western New York through the following comprehensive programs: leadership training, sex education, pre-school guided observation, parent education, and counselling services.

7. A Supplementary Educational Program (North Plainfield, N. J.) The major objectives of the program will be to assist immature children to improve gross and fine motor coordination and perceptual and sensory organization, to relate more comfortably to other children

and adults, and to develop a greater sense of competency about themselves. The program will involve the services and facilities of the Y.W.C.A. and representatives of the staff of Rutgers University.

8. Human Relations Laboratory Training Innovation in Curriculum (Springport, Mich.) The project is designed to plan and develop a human relations laboratory experience as a part of a high school curriculum designed for teaching interactive skills and for improving the learning climate of the high school.

9. Neighborhood Centers for Mental Health and Career Planning (Los Angeles, Cal.) The purpose of the project is to provide, under school sponsorship, mental health, counselling and career planning in non-school buildings in or near neighborhood shopping centers. These services will be available to all school-age children and adults in the community on a walk-in basis.

10. Berkshire Educational Co-operative (Sheffield, Mass.) The purpose of the project is to provide and plan for, on a regional basis, those co-operative educational services and facilities which none of the 13 rural school systems can independently justify or support and to develop innovative educational programs in the arts and sciences which utilize the particular unique resources of the Berkshire area. Some of the activities will include special classes and programs for the mentally retarded and emotionally disturbed children, a diagnostic and remedial reading clinic, an inservice teacher training program,

curriculum materials centers, and the utilization of local cultural resources.

11. Finger Lakes Region Supplementary Education Service Center (Auburn, N. Y.) An eleven-county area will provide a wide variety of cultural experiences in American life, Indian love, foreign cultures, and the arts and sciences, as well as classes in dance and the arts. The present program will be expanded by operating mobile units and by increasing the basic, mobile program with new units.

Operational Grants: (In-School Activities)

Five of the 19 proposals selected at random from those submitted to the reviewer were for operational grants to improve the quality of the educational experiences within the schools. Areas of emphasis include very early childhood education, an interdisciplinary approach to teaching reading skills, a non-graded school organization, a visually oriented primary reading curriculum, and a program of early diagnosis and correction of learning handicaps in the primary grades. A brief description of each of these five proposals follows:

1. A Supplementary Center for Early Childhood Education (Englewood, N. J.) The program includes reorganizing the school system in the early childhood years into a non-graded structure beginning at age 4, demonstrating one way to consolidate the gains of pre-school education with special emphasis on developing those elements and factors which shape the way in which a child learns to learn. The center

will translate research into classrooms and school practice, particularly in the fields of child development, curriculum materials, methods and techniques of instruction, and administrative organization.

2. An Interdisciplinary Approach to Teaching Reading Skills (Fort Dodge, Iowa) The project will guide and direct a program whereby reading skills will be taught within the content areas of science, mathematics, social studies, and English.

3. The Development of a Unique Learning Center for the Implementation of a Student-Centered Instructional Program (Lexington, N. C.) The program will focus on the development of a non-graded school in which efforts will be directed toward the integration of all subject areas into a multi-phased curriculum which will make wide use of resource and instructional media to stimulate in students creative expression, critical thinking, imagination, and desirable mental attitudes.

4. Project Appolo: A Research Project to Develop a Visually Oriented Primary Reading Curriculum (West Sacramento, Cal.) The program will endeavor to develop a visually oriented primary reading and spelling curriculum which will aid that one-third of all children in the general population who read and spell below the level of their ability as measured by standardized tests. Selection procedures will cluster children with similar visual needs into groups of appropriate size for efficient instruction. Moreover, special teaching techniques

and curriculum materials will be developed, evaluated, and disseminated.

5. A Supplementary Program in Perceptual Training for Readiness for Academic Work in the Primary Grades (Perth Amboy, N. J.) The program will attempt to prevent future failures in academic skills, particularly in reading, through an early diagnosis of learning handicaps (in kindergarten) and through working intensively with approximately one-fourth of all the kindergarten pupils (125) whose handicaps to learning are found (through screening) to be most serious. Hopefully, those handicaps to learning will be removed before the pupils fail in school, thus enabling them to find school rewarding rather than frustrating. A team directed by a psychologist-coordinator will consist of five teachers, one social worker, one speech correctionist, one optometrist trainer, one secretary, and screening optometrists as needed.

Analysis

The review of the evaluation section of the 19 randomly selected proposals reflected, in general, below average concern according to PACE standards, for incorporating evaluative plans in proposal applications. On the five-point scale that was used to analyze the evaluation section more proposals (29 percent) received the lowest rating than any other rating. The remaining proposals, as shown in the table on the following page, were rather even divided among the other points on the

COMPARISON OF RATINGS OF THE EVALUATION SECTION OF OPERATIONAL GRANTS

	Ratings									
	1		2		3		4		5	
	Score	Per Cent	Score	Per Cent	Score	Per Cent	Score	Per Cent	Score	Per Cent
I. School-Community										
A. General Description										
Before 1967	11	44	10	40	4	16	..	..	..	..
In 1967	11	37	5	17	6	20	4	13	4	13
Total	22	40	15	27	10	18	4	7	4	7
B. Approaches										
Before 1967	23	46	11	22	8	16	7	14	1	2
In 1967	22	37	10	17	7	12	12	20	9	15
Total	45	41	21	19	15	14	19	17	10	9
C. Techniques Selection										
Before 1967	11	73	1	7	3	20	..	..	..	..
In 1967	5	25	7	35	4	20	..	..	4	20
Total	16	46	8	23	7	20	..	..	4	11
Application Procedures										
Before 1967	11	73	1	7	3	20	..	..	..	..
In 1967	2	10	10	50	4	20	1	5	3	15
Total	13	37	11	31	7	20	1	3	3	9
D. All Areas										
Before 1967	56	53	23	22	18	17	7	7	1	1
In 1967	40	31	32	25	21	16	17	13	20	15
Total	96	41	55	23	39	17	24	10	21	9

	Ratings									
	1		2		3		4		5	
	Score	Per Cent	Score	Per Cent	Score	Per Cent	Score	Per Cent	Score	Per Cent
II. In-School Activities										
A. General Description										
Before 1967	1	10	..	..	1	10	8	80	..	..
In 1967	..	..	2	13	4	27	4	27	5	33
Total	1	4	2	8	5	20	12	48	5	20
B. Approaches										
Before 1967	1	5	1	5	2	10	8	40	8	40
In 1967	3	10	1	3	4	13	6	20	16	53
Total	4	8	2	4	6	12	14	28	24	48
C. Techniques Selection										
Before 1967	1	8	4	33	2	17	3	25	2	17
In 1967	..	..	..	..	4	29	5	36	5	36
Total	1	4	4	15	6	23	8	31	7	27
Application Procedures										
Before 1967	1	8	6	50	2	17	3	25	..	..
In 1967	..	..	..	..	4	29	5	36	5	36
Total	1	4	6	23	6	23	8	31	5	19
D. All Areas										
Before 1967	4	7	11	20	7	13	22	41	10	19
In 1967	3	4	3	4	16	22	20	27	31	43
Total	7	6	14	11	23	18	42	33	41	32

	Ratings									
	1		2		3		4		5	
	Score	Per Cent	Score	Per Cent	Score	Per Cent	Score	Per Cent	Score	Per Cent
III. All Grants										
A. General Description	23	29	17	21	15	19	16	20	9	11
B. Approaches	49	31	23	14	21	13	33	21	34	21
C. Techniques Selection	17	28	12	20	13	21	8	13	11	18
Application Procedures	14	23	17	28	13	21	9	15	8	13
D. All Areas	103	29	69	19	62	17	66	18	62	17

scale: 19 percent received a rating of 2; 17 percent, a rating of 3; 18 percent, 4; and 17 percent, 5. Similar results were obtained for the main areas included in the analysis. For example, more proposals received a rating of 1, the lowest rating, in general description (29 percent), approaches (31 percent), and selection of techniques (28 percent) than any other rating. For application of techniques, more proposals (28 percent) received a rating of 2 than any other rating.

Evidence that evaluation had been considered during the development of the three planning proposals was small, if not entirely omitted. It was sketchy in one planning proposal and, for all practical purposes, almost worthless for informing a reader of the extent to which it had been considered for providing guidance during the planning phase. In another proposal there was evidence that evaluation had been considered concerning the feasibility of the program and the desire to focus the objectives during planning on designing an operational program which would produce measurable or observable qualitative changes in people. Comprehensive plans are not appropriate in a planning proposal; but at least the plans should reflect a concern for some of the main features of the evaluative process, recognizing the contributions which evluation can provide.

The evaluation sections of the 16 randomly sampled grants reflect varying degrees of familiarity with, and concern for, evaluation. However, there is a marked difference in the adequacy of the designs of the evaluative process in the operational grants involving school-community relations and in the designs of those involving the more familiar in-school activities. The designs of the former group are very adequate as compared to those of the latter group. Approximately two-thirds of the proposals concerned with school-community relations received a below average rating: 41 percent received a rating of 1; and 23 percent, a rating of 2. On the other hand, approximately two-thirds of the proposals concerned with in-school activities received an above average rating: 33 percent received a rating of 4; and 32 percent, a rating of 5.

The superiority of the designs of the evaluation sections in the in-school activities proposals over the school-community relations proposals is reflected in each of the four main areas of the analysis: general description, approaches, selection techniques, and application procedures. Approximately forty percent of the school-community proposals received a rating of 1 in each of these areas, while only four to eight percent of the in-school proposals received such a rating. In sharp contrast, 20 to 48 percent of the in-school proposals received a rating of 5 in each of these areas.

The increased emphasis given to the evaluative process, as evidenced in the successive editions of PACE, is reflected in the designs of the evaluation sections of the sampled proposals. More proposals approved in 1967 received an above average rating than those approved prior to 1967. The above-average in-school proposals increased from 60 percent to 74 percent. In general, the improvement was in each of the four areas: description, approaches, techniques, and procedures.

The following summary of the evaluation sections contained in the sampled proposals concerning school-community relations is appropriate.

A. General Description

1. The plans are not clearly and comprehensively presented. The objectives are stated in very general terms, such as to develop understanding of, to see the consequences of, to utilize all community resources, and to investigate and co-ordinate certain services. Relationships between the stated objectives and the brief, generalized evaluative procedures are difficult to perceive. A review of an entire proposal is generally required to read between the lines and imagine the types of procedures which may be used to evaluate the program.

2. About half of the plans contain some evidence that evaluation is considered as a function of operation.

3. In general there are no innovative aspects included in the evaluative procedures.

4. Only about half of the proposals have budgetary commitments commensurate with the proposed design.

B. Approaches

1. About three-fourths of the proposals show evidence of concern for the feasibility of the program: appropriateness for meeting identified needs, compatibility with existing programs, timeliness, acceptance with support, and availability of staff and other supportive resources.

2. Half of the plans state or imply that the design of the program focuses on measuring or observing changes in people, rather than in things: changes in attitudes, cooperativeness, and understanding; improved adjustment to school and home life; and effects of additional cultural and social experiences.

3. Approximately half of the plans do not show feedback during the operational stage, needed to correct deviations and to implement newly identified, promising practices. Also, an equal number are silent concerning the use of evaluative procedures to provide a basis for future programs.

4. Most of the plans do not identify norms or standards to be attained and variables to be evaluated.

5. Approximately two-thirds of the proposals do not give sufficient information concerning techniques to enable one to determine whether the specified techniques are appropriate for measuring the degree the objectives have been obtained.

C. Techniques

1. Outside consultants and standardized tests will be used more extensively during evaluation than other techniques. Only twenty percent of local educational agencies have identified the types of consultative services or the individuals to be used. The manner in which the services will be utilized has been identified by only ten percent of the agencies. The most frequently made statement concerning the use of consultants indicated that outside consultants will be sought either to conduct the evaluation or to give guidance. About twenty percent of the local agencies had sufficiently specified the types or kinds of standardized tests which will be used and the general procedures for using them.

2. About one-third of the agencies stated that pre-post testings, student questionnaires, and extent of participation will be used in the evaluation of the programs. Selection and application procedures will be formulated during the operational stage.

3. About half of the plans only stated the techniques that will be used but did not provide sufficient information to determine the appropriateness of their selection or application.

4. On an average, a combination of three techniques will be used in the evaluative process. The frequencies of use are:

Technique	Frequency
Standardized tests	5
Pre-post testing	4
Teacher reaction	
Interviews	1
Questionnaires	3
Student reaction	
Interviews	3
Questionnaires	4
Project director's reaction	
Case studies	1
Questionnaires	2
Outside consultants	6
Parents' reactions	1
Increased participation	4
School administrators' reactions	1

Interpretation

The following interpretation of the analysis of the evaluation section of the sampled, school-community proposals is based on the belief that an evaluative process should provide feedback during the operational stage at a time and in a form that can be useful in furthering the accomplishment of the objectives, be based on the performance of the participants, actively involve staff and participants from the beginning to the end of a project, be considered as a decision-making tool, and be theoretically based.

1. Planners need additional assistance in designing appropriate evaluative procedures. The findings of this review reveal that planners can more appropriately design procedures for evaluating the more familiar, in-school activities than the less familiar, school-community relations. It is quite possible that causative factors for this difference

are related to differences in the extent of past research and available theoretical knowledge concerning these two areas.

2. Planners need additional time to formalize evaluative procedures. The greatest portion of planning time is devoted to the development of the essence of the program. This traditionally based distribution of time should not pose as a threat to planners or administrators of programs in regard to their being successful in having programs funded initially or on a continuation basis.

3. In general, the evaluative procedures as outlined in the sampled proposals have little value as guidelines for determining the effectiveness of the various programs. Many proposals imply an effort only to meet stated requirements. Restrictive and inflexible plans are not sought. Arguments can quite well be presented to justify submission of evaluative plans after programs have been in operation for a period of time.

4. The inadequacies of the various plans do not reflect the thrust and tremendous promise that the various programs have for producing significant changes in school-community relations. By reading entire proposals, one gleans the sincerity, desire, and determination that the various representatives of local educational agencies have for improving their respective educational programs. The will to succeed on the part of the local representatives and the faith on the part of the funding agencies that the local representatives

will succeed in promoting successful operations.

Recommendations

1. CONDUCT SMALL SEMINARS TO ASSIST LOCAL AGENCIES, WHICH ARE OPERATING PROGRAMS WITH SIMILAR OBJECTIVES IN THE DEVELOPMENT AND REFINEMENT OF THEIR EVALUATIVE PLANS.
2. IDENTIFY TEAMS OF CONSULTANTS WHO HAVE COLLECTIVE EXPERTISE IN THEORY AND PRACTICE RELATED TO SCHOOL-COMMUNITY RELATIONS. ACQUAINT LOCAL PROJECT DIRECTORS WITH THE IDENTITIES AND CAPABILITIES OF THE TEAMS.
3. CONSIDER LESSENING THE EMPHASIS ON SUBMITTING EVALUATIVE PLANS BEFORE A PROGRAM HAS BEEN IN OPERATION FOR A PERIOD OF TIME. THIS WILL NOT DIMINISH THE IMPORTANCE FOR THE PLANNERS AND ADMINISTRATORS TO BE FULLY AWARE OF THE SIGNIFICANCE AND POTENTIAL CONTRIBUTIONS OF EVALUATIVE PROCEDURES WHICH ARE OPERATIONALLY AND THEORETICALLY BASED.

THE GIFTED AND THE DISADVANTAGED[*/]

The sampling of more than 40 proposals approved during the first year of Title III led me to observe that "concern with evaluation is minimal in both the planning and operational grants dealing with the gifted and the disadvantaged." The guidelines at that time did not require proposals to include evaluation and few projects did. PACE guidelines issued since mid-1966, do require that evaluation methods, techniques and procedures be described "where applicable" and that "a pilot or operational activity or both, must include a description of provisions for evaluating the project if it is approved."

Of the 35 proposals in the present sample (projects approved in mid-1966 and early 1967), 17 are for planning grants and another 12 are for both planning and operation grants. Thus, all but six proposals involved planning of programs. The proposal initiators of planning activities tended to ignore the evaluation guidelines and failed to perceive it as an integral part of the process. This observation about planning grant proposals is a reflection of the "state of the art" since it applies to evaluation generally.

As interpreted by those who prepared the proposals in the sample, evaluation is equivalent to data gathering—with the more data available, the better, regardless of its relevance or meaning. Even pre- and post-project assessments are not always proposed for

---

[*/] Prepared by A. Harry Passow, professor of education, Teachers College, Columbia University.

operational grants, often, reactions are gathered or counts made at the completion of the project only. Evaluation is viewed as judgmental— how effective was the project? What changes were brought about? — and seldom is perceived as contributing to the decision-making process. Were evaluation viewed as a factor in decision-making, it is more likely that proposals for planning grants would involve appropriate procedures for assessment and appraisal as part of that process, determining needs and establishing priorities.

Procedures and Approaches to Evaluation

The approaches to evaluation are quite varied, ranging from the use of a single standardized test of reading to relatively sophisticated techniques of interaction analyses. One proposal states, "If the developed programs are feasible, then our objectives will have been achieved." However, no details are set forth for determining feasibility. Another proposal restates the program objectives in the form of the query, "To what degree did it. . . ?" Again, the procedures for determining the extent to which the program may have "strengthened the teacher's knowledge and control of subject matter in the communication skills of speaking and interacting" were not detailed. A third proposal simply promises to select some consultants who will be asked "to work up what constitutes a good program." Still another proposal has a lengthy list of assessment procedures which appear to have come from a textbook on measurement and evaluation.

However, a concluding sentence indicates vaguely that "appropriate instruments and procedures will be selected from the above listing."

The procedures and techniques referred to in the various proposals included the following:

Standardizes tests—intelligence (group and individual), achievement (especially reading and mathematics), aptitude, personality, and diagnostic (e.g., reading difficulties).
Teacher or locally prepared tests of achievement.
Ancedotal records by teachers, counselors or outside observers.
Check lists.
Inventories (particularly of interest).
Rating scales—self, teacher, parent, observer, counselor, etc.
Questionnaires.
Sociometrics.
Film, tape and videotape recordings.
Interviews.
Observations.
Reactionnaires and testimonials.
Participation surveys—attendance, usage, retention, etc.
Interaction process analyses.
Adjustment ratings.
Self-concept assessment instruments.
Group discussion.
Follow-up studies.
Written summaries.
Sociological surveys.
Written essays.
Physical examinations (i.e., vision, hearing, etc.)
Staff reports.

The above list is relatively inclusive and, while compiled from the proposals in the sample, could have been drawn from other sources such as the Guide to Evaluation of Title I Projects or A Guide to Assessment and Evaluation Procedures or any basic textbook. One implication of such a listing is that the USOE guidelines may well be a useful vehicle for sensitizing school practitioners to evaluative techniques other than commercial standardized tests.

Analysis of Evaluation Procedures and Approaches

Compared with the first year PACE proposals dealing with the gifted and the disadvantaged, the current sampling indicates some greater awareness of assessment procedures, instruments and techniques. However, awareness alone is not enough to represent real "progress."

The two most sophisticated evaluation designs were prepared by university consultants and staff members for consortia of schools. A proposal for a Supplementary Educational Center for a five-county area in a western mountain and desert region took each of the three "pupil change objectives" and the five objectives directed at the "overall improvement of the educational program" and indicated the means by which assessment would be carried out—both in the project schools and in a carefully selected control school. This proposal included illustrations of how data were to be analyzed. A proposal for using Indian heritage and culture to build an adequate self-image and improve English as a second language, related assessment and data-gathering procedures to the three specified goals of the program. In addition to pre- and post-data, the proposal detailed periodic ongoing observations and recordings, tapes of students' speaking, reading and participating in discussions; anecdotal records of pupil behavior in dormitories as well as classes; self-image rating scales and essays; adult observations; etc.

In general, the concept shaping evaluation proposals has four stages: gathering some pre- and post-data, looking for changes in scores, assuming that whatever differences do occur are significant, and attributing these to whatever services or treatments were applied. Existing standardized instruments are used with the assumption that their "objectivity" is valuable, regardless of their suitability. If the practitioner is aware of the limitations of the standardized tests with respect to his own population and program, this is not revealed in the proposal. More subjective procedures ignore essential modification to specific programs and populations or the time requirements in applying and analyzing such techniques. Just as many commercially published achievement tests fall outside a particular program, rating scales, inventories, personality assessment instruments and other devices usually require considerable adjustment by a skilled technician to local needs. Analyses of tape recordings or videotapes, for example, are costly and time consuming (as are most process studies) and yet, few proposals recognize the budget demands implied. Except for money for test purchases, funds for an assistant and some consultative service, and, in some instances, a sum for computer time, most budgets peg costs of evaluation too low. In few instances are illustrations given of the kind of items which might be included on a questionnaire or in an interview, the listing of the generic type of device or procedure is considered sufficient.

In summary, proposal writers have generally responded to the requirement for describing evaluation provisions by listing instruments and devices from which a selection will be made "later" and institutions or individuals who will be asked eventually to assess project activities. The impression one gains from reading the proposals is that, with few exceptions, the evaluation section is written hopefully to satisfy the Office of Education's requirements—not because evaluation is perceived as having any intrinsic value for the program itself. Thus, there are two complementary jobs to be done: (1) prepare materials which will convince the proposal writer that evaluation is really important to the program; exhortation will not achieve this and (2) prepare guidelines which will provide him with the basic instruction he needs to prepare meaningful evaluation designs.

Some Interpretive Observations

To repeat, the confusion and lack of clarity concerning the purposes and nature of educational evaluation in general, and with respect to PACE in particular, are reflected in the proposals. A number of questions recur. To what extent are the purposes of local evaluation congruent with those of the Office of Education's need for "hard data" for Congress? Do local project directors require the same kinds of information for their purposes as the Office of Education does for its? Since procedures which describe changes are not necessarily those which will explain or account for those changes, how "tight" should the evaluative design be? Is it sufficient to describe growth and indicate associations or must PACE projects determine cause-and-effect relationships? (Obviously, since Title III projects are not research programs and are intended to be innovative and exemplary, they cannot be structured so as to relate specific inputs to particular outcomes.) How creative should evaluative designs be for supposedly innovative projects? In fact, most of the proposals lack any semblance of an evaluation design—instruments and personnel are considered the design, rather than its elements. Assumptions and conditions are inadequately described and populations often lack specificity.

PACE projects dealing with the disadvantaged (four-fifths of the 35 proposals in the sample were in this category) tend to overlap with Title I, ESEA which is aimed at improving education for the same population. To what extent can the Guide to Evaluation of Title I Projects (October 1966) be helpful for proposals which are being prepared for Title III dealing with the disadvantaged? The evaluation process which Neidt and French describe in that guide—"the process of determining the extent to which specified objectives have been reached" or "the process of assessing the extent and direction of change resulting from an educational experience"—is sufficiently generalized to be applied to activities of many kinds other than those dealing with the disadvantaged. In fact, were not the examples drawn from programs for the disadvantaged, the document could serve as a concise primer on educational measurement and evaluation, complete with basic statistical techniques.

The focus of the Title I guide and the evaluation designs it presents is on outcome or product evaluation: "Evaluation designs are simply procedures which allow the experimenter to derive meaning from the amount and direction of changes which have occurred in a project group." The six types of designs which Neidt and French use as illustrative, all involve comparisons of data derived from within the project group itself or from an outside population. They do conclude that it is important to evaluate continuously throughout the Title I project in addition to assessing the final outcomes, but do not provide detailed illustrations or discussion as to how such periodic and ongoing assessments should be made.

Since evaluation is concerned with "effectiveness," most designs tend to be limited to anything other than overall impact. "Net impact" evaluation may be all that is possible for PACE projects for the kind of data required and the designs needed for determining which of several factors (services, programs, conditions, etc.) is responsible for any changes are difficult to obtain.

An encouraging sign in some proposals is an apparent awareness that both intellective and non-intellective factors affect change. In addition to the usual circumscribed measures of academic achievement, techniques are being proposed for assessing other factors in the cognitive and affective realms. For example, ratings of self-concept and perception of self are frequently mentioned in programs for the disadvantaged.

Still, evaluation continues to be viewed as an appendage to a proposal, almost an afterthought. Hammond has observed:

> Research has failed to produce adequate guidelines and procedures to be utilized by school districts for the purpose of evaluating both current and innovative programs. The problem is complicated further by the fact that the school districts of the past have not included the process of evaluation as one of the major criteria for curriculum improvement.[1]

Far more needs to be done in developing and using evaluation totake continuous readings of program developments in order to make better decisions along the way. Stufflebeam's definition of evaluation "as the process of acquiring and using information for making decisions associated with planning, programming, implementing and recycling activities" seems particularly appropriate for PACE activities. The CIPP model (context, input, process and product) seems promising although far more needs to be done to communicate its basic dimensions to relatively unsophisticated proposal writers. The notion of "context evaluation" which deals with defining "the environment where change is to occur, the environment's unmet needs, and the problems underlying those needs" seems to have particular relevance for individuals preparing planning grants, for example.[2/]

---

[1] Hammond, Robert L., "Evaluatation at the Local Level." Tucson, Arizona: PROJECT EPIC. p. 2. mimeographed, undated.

Recommendations

One is tempted to recommend that innovative and exemplary programs be required to include innovative and exemplary evaluative designs. However, a great deal is yet to be done which would enable better determinations of "the extent to which funds provided under this title have been effective in improving the educational opportunities of persons in the area served. . . ." It is recommended that:

1. CLARIFY THE PURPOSES OF EVALUATION AS THEY SPECIFICALLY RELATE TO PACE PROJECTS WITH RESPECT TO LOCAL AND NATIONAL NEEDS.

While there is obviously some overlap in the kinds of data required for determining effectiveness, there are also differences. What the local educator requires for decision-making as he develops and implements a PACE project may be of a different order from that which the Office of Education needs to determine how effective Title III is as a whole. The relevance of certain kinds of data and evaluation designs will differ as the objectives vary. The fuzziness with which this issue has been dealt with to date has not contributed to sharpening the evaluation needs at local, state and national levels.

[2/] Stufflebeam, Daniel L., "The Use and Abuse of Evaluation in Title III." Address delivered at the National Seminar on Innovation, Honolulu, Hawaii, July 1967, p. 8.

2. PREPARE A GUIDE TO EVALUATION OF TITLE III PROJECTS WHICH CONTAINS THE BASIC INFORMATION NEEDED TO DEVELOP EVALUATION DESIGNS APPROPRIATE TO PACE.

Such a guide has been developed for Title I, ESEA, and experience with its use has been accumulating. That guide is general enough to cover a variety of designs, mostly in the traditional experimental sense. The CIPP Model and the PROJECT EPIC Model (as described in the Stufflebeam and Hammond papers) are too terse for the educational practitioner in the present form. However, both models could be expanded with considerable illustration, and serve as the core of a guide for evaluating Title III projects. With hundreds of projects having been funded—planning, operational and pilot—surely there must be some which are innovative and exemplary with respect to their evaluative dimension. These might be identified and presented as illustrative for guiding the design of evaluation programs. Finally, if the Office of Education requires specific kinds of data for its purposes, these should be specified in the Guidelines.

3 RESERVE SOME PACE FUNDING FOR INNOVATIVE AND EXEMPLARY PROPOSALS DEALING WITH EVALUATION.

These could be "invitational," involving a center with evaluation design capabilities and a basically sound PACE proposal, with the former helping the latter to develop and implement the evaluation of the project. The description and reporting of such joint ventures could be disseminated to all groups currently receiving PACE information and be incorporated in the guide suggested above.

Effective evaluation requires skilled, technical competence. There is a shortage in personnel with such trained competence. Detailed guides are of some help but they are generally inadequate alone. This is evident when one reads proposals and finds the writer rephrasing the sentences of the guidelines without conveying understanding. A final recommendation would be aimed at building liaison with other federally-supported programs for technical help in designing evaluation programs for PACE. The Research and Development Centers, the Regional Educational Laboratories, and some of the research training programs might conceivably help develop PACE projects without impairing their own missions.

CLASSROOM PERSPECTIVE[*/]

Twenty proposals were examined to see what devices were being used to evaluate the project objectives. All proposals contained a separate evaluation section—some lengthy, others short. Common procedures involving tests, questionnaires, group discussion, and interviews with individuals were found, but there was also evidence of a few unique and original instruments to provide feedback.

A few proposals indicated an expert, or college or university would be given the evaluation responsibility. In one case the expert's name was mentioned. However, nothing was shown as to how either facility might proceed.

One evaluation section contained the following statement: "Title III of E. S. E. A. is intended more to stimulate innovation than to finance extensive evaluation of program results." Listed below is a random sampling of stated evaluative measures:

A. "Objective evaluation as made with the Iowa Test of Basic Skills and the Iowa Test of Educational Development. Scores will be compared to local and national norms to chart student success."

B. "The Research-Planning-Development Division will be specifically charged with devising suitable evaluative techniques and instruments. Working closely with research

---

[*/] Prepared by Joseph B. Rubin, teacher, Chapman Elementary School, Portland, Oregon.

agencies, this group will conduct a continuous evaluation program which will provide a feedback for constant improvement." "Comparisons will be made of reading readiness test scores obtained for kindergarten pupils of previous years."

C. "Interest and success of students."

D. "Pupil and teacher adaption and progress made according to the various materials and instructional techniques utilized within the center."

E. "Evaluation strategies for use by involved district personnel will be developed during the planning phase of the project."

F. "Follow-up questionnaire will be used to determine how many changes were made by visitors to the program. From there, it will be determined how successful the program is in terms of teacher behavior."

G. "Increased library circulation, compared to classes not in the study and not counting books in the rooms. Involves not only quantity but quality." (Base: Records over a four-year period.)

H. "Authorities from universities will be used in developing appropriate evaluation tools and techniques."

I. "The administration of appropriate surveys and/or tests recommended by the State Reading Supervisor, Mr. John Doe, to all participating teachers prior to and at the conclusion of the program."

J. "As the field of independent study is so innovative there do not exist many academically accepted means of evaluation; therefore it has been decided to engage a graduate student who will design an acceptable evaluation program for independent study in literature as his doctoral thesis."

K. "Opinion of coordinator, based on as much objectivity as possible."

Some of the evaluation sections included "all kinds of ways" to go about measuring learning. These ways did not occur in any particular order or sequence.

Most of the evaluations were concerned with acquisition of information, scholastic achievement, and/or conformity to predetermined goals. Only a few of the proposals included evaluations showing concern for change—(change that had come about in concept of self, ability to cope, enlarged horizons, or creativeness on the part of pupils and/or teachers). There was recognition of kinds of growth not programmed into the original proposal.

Most proposals gave a choice for decision-making on the part of the evaluator, while other sections of the proposals were usually rigid with no apparent choices for anyone.

Two of the 20 proposals studied included specific procedures for assessing each stated objective of the project. This consultant found, however, that most evaluation activities were not related to the specific objectives of their project.

Evaluations fell into two categories: (1) evaluations that went on continuously as part of the instructional process, and (2) evaluations that covered the total program or project. While the two cannot be completely separated, each demands a different kind of assessment. Most programs focused on either one or the other of these, few on both.

Interpretation

Most evaluation sections of these 20 examined projects were insufficient. Poorly stated objectives, lack of imagination, and not knowing "how to assess" much of what had happened are the reasons for my observation. In no way has Title III imposed restrictions upon these sections. The similarities amongst techniques flow from a feeling that established procedures are always valid and are appropriate to measure outcome.

There were some specific activities designed to see what happened to the individual, not only to the group. This was admirable for too many projects were aiming at stereotyped responses. On-the-spot reactions, learner's feelings, attitudes toward goals were not considered. Certainly, learning and how it takes place were viewed differently by each Title III project.

Recommendations

To improve and make more sufficient the evaluation sections of Title III proposals, it is recommended that the following beliefs be considered vital to evaluation techniques.

1. EVALUATION MUST NOT BE CONFUSED WITH TESTING.

The two must be clearly distinguished. Testing is a contrived situation resulting with a pencil response to something acquired. The purposes of evaluation are: to identify learning stages recognizing progress, to re-establish or alter the learner's purposes, and to plan for the next progressions. One can easily see that evaluation is useful as feedback and as a guide for further learning.

For what is learned is reflected in real situations and the evidence can be seen.

Because one has acquired, does not necessarily mean his behavior changes. Examples: a driver may know how to handle a car correctly and pass a test 100%. But, he may also drive in no way that he knows is lawful. A child may spell all his words correctly on a test but write them incorrectly in an original story.

Most test scores are of little value. The only valid proof of learning is evidence that a person's behavior becomes different if and when situations become pertinent.

2. ABILITY TO EVALUATE ONESELF IS A SKILL OR ATTITUDE WHICH CAN BE LEARNED.

This type of assessment is mainly concerned with that which is valuable in itself—the effectiveness of dealing with others. In includes more than possessing facts.

A transference of responsibility must occur. To function in this way brings to the learner the responsibility for what is being done and how it is done. The entire process of evaluation, so complex and difficult, becomes more meaningful to those who seek a behavior change and are committed to this goal. These skills of assessing one's attitude and abilities increase growth opportunities.

Title III can provide some guidelines to ensure more adequate evaluation procedures in future proposals. It is the opinion of ths consultant that the above two principles would be useful.

COMMUNICATIONS[*/]

Eighteen of the approved proposals for operational grants have been read for this assignment. Twelve of them were submitted before July, 1966; the others between that date and April, 1967. They represent among $6 million on education expenditure. In one way or another each of them is a program for disseminating educational information or new educational practice. For example, they include proposals for a bus to serve as a traveling classroom, carrying pupils on field trips; several mobile vans to carry materials or demonstrations to outlying schools; a demonstration school; several audio-visual or learning resource centers; and a general program to encourage innovation in teaching methods. The numbers are insufficient, of course, to establish a reliable trend line, but I was unable to detect a significant trend, anyway. Therefore, I am going to discuss the proposals together, rather than in terms of earlier and later styles.

Some of the plans were quite imaginative, and there was no close correlation between size and imagination. They showed evidence of work—particularly, in efforts to consult a large number of persons before drawing up the proposal, in obtaining endorsements, and in attention to such details as specifying the equipment to be bought. Indeed, there was sometimes more attention to equipment than to objectives or personnel. But the proposals scrupulously followed the

---

[*/] Prepared by Wilbur Schramm, director, Institute for Communication Research, Stanford University.

form provided for in the manual, and in so doing were careful to include a section on evaluation. These sections were correctly placed and numbered. They followed the scheme, often the very words, of the instructions for making proposals.

And I must say that they were most marvelously unimpressive!

Kinds of Evaluation

What kind of evaluation was proposed? Almost wholly a study of outcomes. Only a few planned to make any use of research at an earlier stage of the project. To be precise, four of 18 gave some indication that they planned to feed research information back into the operation while procedures were still changeable. Only one specified a pre-operational study—in this case, a survey of facilities and equipment. But each one specified some measure of effects.

What kind of measures? The most common one consisted of collecting records of the use of the project—number of visitors, circulation of materials, enrollment in courses, and the like. Twelve out of 18 were going to do this sort of thing. Eleven were going to collect opinions—from users, teachers, parents, school administrators, and so forth—as to how well the project had worked. Three called for self-evaluation on the part of the teachers or pupils who used the project services—did they feel the project had been good for them? Three called for experts to observe the users of the project, and three others for outside evaluators to come in and evaluate—somehow. Six provided for performance tests of some kind.

What kind of success were they trying to measure? The sharp thinking about success tended to stop with the point where information had reached the target. They were generally quite clear about the kind of user data they could collect, rather vague about the kind of behavior they hoped to accomplish in the user. Typically the proposals said that they:

> . . . will establish. . .
>
> . . . will operate. . .
>
> . . . will provide. . .

One of them couched the objectives in terms of "models":

> . . . will devise and operate a procedural model. . . etc.

That is to say, they were chiefly concerned, as most communicators are, with delivering the message. Beyond that? Well, they would collect opinions. They would ask the guidance department of a nearby junior college to select achievement tests (no discussion of how to separate out project effects from other effects!). They would hire a consultant. They would have a "jury of professional evaluators". They would measure the "performance of children with use of various educational media and related to objectives of Title III and in relation to purposes of Title I".

One comes away from reading these evaluation sections rather certain that (a) evaluation means very little in these project plans; (b) little thought has been given to what constitutes success—at least behaviorally—in these projects, and (c) few of them have had the benefit of much attention from a research man.

The evaluations are not integrated into the projects. In most cases, little will depend on the measurements; they seem to be in the plan mostly because the Office of Education wants them there. Almost no research personnel is listed among the project staffs. In general, the work is going to be turned over to a consultant, or made "a responsibility of the director." Only one full-time research officer is listed among the 18 projects. Five consultants are called for, though not named. Two visiting teams of evaluators are mentioned, rather vaguely. Help from neighboring institutions is several times called for. But in general, evaluation is going to depend upon records of the use of the project, to be assembled by the director or one of his staff, or someone outside the project is going to be called upon to make some observations or measurements.

Furthermore, evaluation is very hard to find in the budgets. Most projects have no line item at all for it. One $1.5 million project specifies $5,000 for evaluation expenses.

In contrast to the lists of materials to be bought, which are detailed and specific, the details on evaluation are vague and general, except for the counts of users and materials circulated. Instead of details on research method, one sometimes feels the wind blowing out of old statistics textbooks—for example,

> "A wide variety of standardized and novel instruments will be employed in a controlled manner. Maximum use of power concepts in statistical treatment, non-perimetric (sic) procedures of data analysis, in addition to more accepted patterns of behavioral science analysis will be constantly employed."

As that salty critic of statistical methods, Quinn McNemar, would have said, this is the kind of headlight that shines backward.

Sentences like the one just quoted, on top of all the other evidence, leave us with the uneasy feeling that a research man hasn't been involved in these proposals, except perhaps perfunctorily, and isn't likely to be; that evaluation tends to be something built in for requirements' sake rather than because it is an integral part of a functioning project; and that the results of any evaluation accomplished as a result of these proposals are not likely to be very soon or very greatly helpful either to these projects or later ones.

Recommendations

Well, what to do about it?

I don't think it will help much to circulate a model. The

Stufflebeam "model" is useful in that it calls attention to the helpfulness and appropriateness of research at every stage of a project. The Stake "model" is useful in that it points out the difference between evaluation of congruences (which chiefly helps the project) and evaluation of contingencies (which is necessary if the evaluation is to contribute much to science or to later users).

But I don't think you will find these models or others very useful in changing the direction of project proposals. Rather, it seems to me, you need to convince proposal writers that:

1. They need the help of a research man.

2. They need him from the beginning of their planning, when they are setting objectives.

3. They need him as a member of the team, where he is likely to be helpful in operation rather than in public relations.

4. They need him at every stage of the project, not merely at the project memorial service.

5. They need to finance his work adequately.

I don't see why this message can't be conveyed in so many words. And you can support it by making available some materials more practically helpful than that section in the manual.

For example, why doesn't the PACE program take a few thousand dollars of its money and make some case studies of projects in which research has been usefully integrated through the course of the operation? The people who are preparing these proposals know very little about research. To most of them it is something esoteric and strange. They see it as a judgment to be handed down by an outside tribunal, or as a device to collect user data for a glowing annual report. The best way to show them that research can be an inside, rather than an outside, tool, that it can be used to answer deep rather than superficial questions, and that it can be most practically useful to the designers and administrators of a project, would be to let them see some of the examples of cases where it has been so used.

A second kind of material to be circulated would be samples of some methods and instruments which have been used successfully at different stages of projects.

Thirdly, as soon as possible, PACE might well undertake some more administrative research of its own, and study a few completed projects to find out what has or has not been accomplished by evaluation. I am afraid the results will be discouraging, but they deserve to be known, and from knowing them we may find out how to encourage better performance.

Now, I realize that what I have been suggesting—bringing research and research men into the project from the beginning—raises serious problems which the prescribed length of this report leaves insufficient space to discuss. Where are you going to get all these research men? You probably can't find enough to ride the great wave of PACE projects. And it may be difficult to find good ones willing to undertake this kind of task. A psychologist—to take the most obvious example—doesn't earn many Brownie points for doing practical evaluation research. This is a basic problem that should concern PACE, and that we should discuss at considerably greater length than is possible here. But I see really only three solutions:

1. We can properly reward some able young men for working a while in this area. This may require a training program which, in itself, would seem to be justified considering the vast amount of educational evaluation that will probably need to be done in the next few decades.

2. We can persuade men whose reputations are already made and whose status is secure, to undertake some evaluation projects. This has happened oftener than one might think, but in itself is not sufficient to meet the needs.

3. We can encourage the combining of evaluation research with the study of basic problems in the same laboratory situation. The classic prototype of this pattern is the program of the Army Information and Education Branch research unit during the war of 1941-45, under the direction of Carl Hovland and Samuel Stouffer. The basic task of this unit was to evaluate the effectiveness of different military films and to answer key questions about the opinions and attitudes of soldiers. These tasks were done brilliantly and the results

were reflected in many policy decisions. But at the same time these imaginative researchers were able to make a basic contribution to military sociology and psychology (The American Soldier) and to lay the basis for the Yale Studies of Communication and Attitude Change, which are responsible for much of modern communication theory. This kind of approach, it seems to me, constitutes our best hope of attracting first-rate men to evaluation research, and I should like to see a few cases made public of instances where it has been successfully done.

What I have suggested will not be easy, but something like it must be done if evaluation is to be more than pro forma in this program.

SUPPLEMENTARY SERVICE CENTERS[*/]

NEEDED - A NEW BREED OF EVALUATORS

At long last, evaluation is "in". The school man has gotten the message. Beset by a bewildering array of how-to-do-it monographs, guides, models and flow charts, the practitioner has "fixed" on this new star on the educational horizon. His initial peek is as furtive as a fugitive's look over the shoulder for the pursuing law. In a sense, he too has been getting away with murder for years, and knows it. The book is being thrown at him—in fact all the books from all sides and at one time.

As in other significant instances of educational change, money has generated thunder and lighting. For decades, voices in the wilderness have cried "Proof! How do you know! Can we count on it! What differences will it make—now and tomorrow!" —to little avail. Researchers charged practitioners with gross indifference to the scientific method and perpetration of fraud on the unsuspecting public. Practitioners, on the other hand, charged researchers with callous disregard for human values, worship of empiricism, and empire building.

Few victories were scored by either side until the passage of the Elementary and Secondary Act of 1965. In an effort to obtain valid data from thousands of projects sponsored under the act's various

---

[*/] Prepared by Ira J. Singer, assistant superintendent in charge of instruction and special services, West Hardford Public Schools, West Hartford, Connecticut.

titles, the USOE has entered the fray as a kind of educational United Nations. Joined by such other emerging allies as regional research associations, educational foundations, research divisions of university centers, state education departments, professional educators' associations, Title IV regional laboratories, and Title III research service centers—the federal government has managed a temporary lull in the hostilities.

An uneasy truce has been declared and wary "peace feelers" are being extended by both sides.

For example, Robert Stake contends that the task of improving the precision of measurement instruments be postponed until an "awareness of a full array of teaching and learning phenomena" has been demonstrated. He states that priority should be given "to a more complete description of what we have observed before we become overly concerned with what we want to discover. We need," Stake says, "a new technology of educational evaluation. . . new paradigms, new methods, and new findings to help the buyer beware, to help the teacher capitalize on new devices, to help the developer create new materials, and to help all of us to understand the changing educational enterprise."[1/]

The need for a fresh new approach to the short-term, action

[1/] Tyler, R., Gagne, R., and Scriven, M., "Perspectives of Curriculum Evaluation," Chicago: Rand McNally and Company, 1967, pp. 2-3.

oriented Title III project evaluation is underscored by a reading of recently approved proposals. A study by the writer of 14 Title III proposals funded as operational supplementary services centers reveals three general approaches to the task of evaluation:

A. A general statement on the importance of evaluation, unaccompanied by any plan relevant to the purposes of the proposal.

B. A general statement of procedures for determining outcomes with no description of objectives, resources, plans and processes.

C. A comprehensive, tightly designed evaluation scheme describing existing resources, behavioral objectives, controlled experiments, procedural strategies, communications devices, measurement instruments, and possible outcomes.

A brief discussion of each approach might serve to assist the reader and potential proposal writer.

A. A general statement on the importance of evaluation, unaccompanied by any plan relevant to the purposes of the proposal. In this category, one operational proposal in outdoor education funded its first year for approximately $110,000 contains the following statement on evaluation as its only plan.

> Special attention will be given to the evaluation of the experiences in outdoor education conducted by the project itself and the classroom teachers. All materials, methods, and procedures will be continually evaluated for effectiveness. Also small scale studies will be made of the effect of the outdoor education experience on the

participating children. Evaluating techniques and instruments developed will be shared for any interested teacher to use with her own programs. Methods, materials and procedures will be revised and modified on the basis of the evaluation process.

That is all - and for $110,000.

Another operational proposal, statewide in scope, includes as its fulfillment of the evaluation requirement the following:

> Provisions will be made to secure the services of a team of evaluators from an institution of higher education. This team will be given specific instructions to evaluate this program in terms of improved education. They will not make suggestions as to how we might improve the program in order that we could follow their lead and thus improve our evaluation rating. This team will not be connected with the project in any manner except as evaluators. The Project Director will carry out a comprehensive evaluation program in terms of participation, observation and opinions of both professional consultants and local professional personnel. These evaluative procedures will be completed as separate results and compared.

The above is worth $157,000.

A third sample of evaluative inadequacy is found in the following plan submitted by a regional services center requesting a first year sum of $143,000:

> As indicated previously, to develop this program, involved were teachers and administrators who were visited individually, not as a group, and completed questionnaires. To get feedback on the degree of success that this program will have, it is intended to involve teachers and administrators from public and nonpublic schools. Personal contact and individual meetings, group meetings and questionnaires will be used with teachers and administrators to evaluate the success of the program. These methods will be used periodically to evaluate the objectives of the "x" unit, the "y" unit, the inservice specialists, kits and community resources and A-V coordinators. Other evaluators will be the "z" consultants used by the program, and other personnel involved in the project.

Once again, a gesture not a plan.

The three projects cited above (21 percent of the proposals read by the writer in preparation for this paper) involve a sum of approximately $410,000 for fiscal year 1967-68. Unfortunately, they do nothing more than confirm the importance of evaluation and inform the reader that something or other will be done by an alien people who know more than the proposal writer but are suspect insofar as their motives are concerned. Completely absent are statements concerning priorities and needs, specific objectives, behaviors, standards of performance, and alternative strategies. Also lacking is a description of existing resources and their potential for helping to achieve the goals of the project. The procedures to be employed in evaluating the processes (monitory variables, detection of deviations, communication to project personnel) and the outcomes (standards of behavior, application of techniques and instruments, analysis of data) are omitted completely. Without this information the readers' guess is as good as the writers' in terms of the significance of these projects.

B. A general statement of procedures for determining outcomes with no description of objectives, resources, plans and processes. Of the 14 projects read, seven cite specific standardized test measures to be used, three indicate the use of questionnaires, six use the interview technique, two mention pre- and post-achievement tests, four claim to use rating scales, three mention controlled experiments, four cite attitude surveys, one lists the daily diary, two call for cost studies, and one mentions simulation techniques. Although all projects indicate teachers and/or students as participants in the evaluation process, only two involve administrators while two others use parents.

Proposal "v" states an intention to measure attitude changes but none of the techniques to be employed are described. Proposal "w" simply lists seven separate standardized achievement tests as its fulfillment of the evaluation function. Proposal "x" offers general statements on proper and improper procedures for evaluation and reveals the writers' intention to obtain pre- and post-standardized achievement test scores from "matched groups" without any description of his reasons for doing so. Proposal "y" lists five different techniques to be employed for gathering data without offering a rationale for any single technique. Proposal "z" incorporates the statement that "an evaluation of the competence of the participants (teachers) will be conducted by the participants, supervisors of participants, consultants, and superintendents." However, there is no prior statement explaining why or how this evaluation is to be conducted.

Four of the proposals in this category do reflect honest doubt about the evaluation procedures described. The writers state that, for their programs, available measures appear inappropriate for assessing success or failure. They are wary of predetermining the course of the project through a premature, airtight evaluative procedure. They are aware of the need for a certain amount of flexibility, unplanned intrusions, and unpredictable outcomes. However, the processes of evaluation are either omitted or scantily treated. Since this group of writers refuses to make predictive statements, it is not surprising that they also omit any discussion of outcomes in terms of data analysis, behavioral change and decision making.

In summary, these latter proposals lack thoughtful rationale and sequential development and depend too heavily upon subjective observation and/or standardized test scores. The dangers of over dependence on standardized instruments are self-evident. Stake points out that "the standardized achievement test is unlikely to encompass the scope or penetrate to the depth of a particular curriculum being evaluated."[2/] While it is entirely proper to gather subjective data and to administer standardized test scores at given times in some experiments, it is inappropriate to turn to standardized instruments as the only reliable measures for eliciting objective information.

C. A comprehensive tightly designed evaluation scheme describing existing resources, behavioral objectives, controlled experiments, procedural strategies, communications devices,

---

[2/] Ibid., p. 6

measurement instruments and possible outcomes. This group of four proposals incorporate evaluation statements containing internal strength. The objectives are well defined, sound rationales are provided, procedures are clearly described and sequentially arranged, and the identifications and applications of evaluative measures to be used are presented in a straight-forward manner.

Despite the evidence of structure, there is the sophistication of flexibility. For example, one proposal writer in discussing the management of the evaluation task states:

> "It is not anticipated. . . that there will be a slavish adherence to PERT management procedures but some modifications which will permit linear task time analysis on other feedback and control procedures which will contribute to the achievement of operational program objective."

Wordy, perhaps, but indicative of predictive thought and planning. The use of a carefully mounted procedural strategy is deemed necessary, but tolerance for sensible deviation is evident.

Other proposals in this group incorporate pre- and post-test and/or survey techniques. Although the subjective interview and questionnaire techniques are common, the use of rating scales, cost analyses, attitude change detection instruments, and simulation techniques are alluded to, albeit in limited detail. Assessments of student responses and teacher behavioral changes are included for measurement against such valid models as Bloom's taxonomy and Taba's teaching strategies. Also common among these proposals is unspoken agreement with Stufflebeam's statement that "anyone who has ever directed an innovative project knows full well that he can't constrain his treatment to its original definition just to insure internally valid end of the year evaluation data."[3/] Such thinking could make the task of evaluation much less self-conscious and more productive.

The need for a realistic field research model is unmistakable. Readers from the United States Office of Education have consistently ranked the criteria relating to evaluation as next to lowest (fourteenth) on the ranking scale.[4/] The poor quality of plans for evaluation are constantly referred to in the special consultants' assessment of the first round of Title III proposals.[5/] Although it is not this writer's purpose to repeat descriptions of several research models analyzed by others in this study, it would be helpful to cite a letter from Kurland to Stufflebeam framing several excellent questions which might be posed to all erstwhile model makers:

1. What sort of information should be produced by a project

[3/] Stufflebeam, Daniel L., "The Use and Abuse of Evaluation in Title III," Ohio State University, Evaluation Center, Columbus, Ohio, mimeographed, July, 1967, p. 7.

[4/] Ibid., p. 3 (Criteria for Dissemination ranked fifteenth).

[5/] Miller, R., ed., "Notes and Working Papers Concerning the Administration of Programs Authorized Under Title III of Public Law 89-10, The Elementary and Secondary Education Act of 1965," Subcommittee on Education, United States Senate, April, 1967.

that would help others know whether it is worth further investigation and adoption?

2. What sort of information should be produced by a project that would help state and federal programs offices decide whether the project ought to be continued?

3. What sort of information should be produced by a project that would help local decision-makers determine whether to seek local support as federal support phases out?

4. What sort of information should be produced by a project that would help develop summary reports on the affects of Title III?[6]

Kurland also suggests some alternative descriptions of Stufflebeam's Context, Input, Process and Product (CIPP) categories as follows:

Context - Stufflebeam - Major objective is to define the environment where change is to occur, the unmet needs in this environment, and the problems which underly those needs.

Kurland - To what problem was the project addressed? In what context did the project operate?

Input - Stufflebeam - Major objective is to inventory relevant capabilities of the proposing agency, strategies which may be appropriate for meeting program goals, and designs which may be appropriate for achieving the various objectives associated with each program goal.

---

[6] Miller, R., ed., "PACE National Study", United States Office of Education, Memorandum No. 5, September 1967.

Kurland - What resources (skill, knowledge, people, facilities, materials) were required?

Process - Stufflebeam - Major objective is to detect or predict, during implementation stages, defects in the design or its implementation.

Kurland - What were the procedures used to achieve the results?

Product - Stufflebeam - Major objective is to relate outcomes to objectives and to context, input and process, i.e., to measure and interpret outcomes.[7]

Kurland - What result was achieved and how did actual outcomes compare with expected?

Common usage of the CIPP model by evaluators across the country could produce valuable information sufficiently standardized for subsequent computer programming and analysis. Such analysis could lead to reduction of duplication of projects, dissemination of experimental results, identification of areas of the overall educational program requiring new thought and activity, and evidence for future local, regional, state and national categorical and general subsidy programs.

Widespread use of the CIPP model will also require the recruitment of large numbers of new personnel for evaluation purposes. In their search for such personnel, Title III project directors usually overlook local professional staff. Yet, if federal, state, and local budgets continue to grow in support of massive local experimentation, the teacher and administration will have to be relied upon more and more. Tyler contends that "through training we could refine the teacher's powers of observation and instruction to make his contribution both technically sound and educationally valid. It is not unreasonable to conjecture that some day the primary role of the classroom teacher may be as. . . an evaluator."[8]

---

[7] Stufflebeam, pp. 9-14.

Stake's plea to involve teachers in responsible monitoring and judgment rendering roles is sheer heresy to the traditional researcher. However, to the beleagured practitioner, the offer brings with it new hope and realism. Stake's comment suggests an end to cultism, a regard for the intellectual prowess of at least a segment of the field worker force, a method by which certain research tasks can be taught to and practiced by teachers, and a desire for dialogue and cooperation between the creators, practitioners, and evaluators of an educational idea. Such an approach might prove Stufflebeam's contention that "evaluation should stimulate rather than stifle dynamic development of programs."[9]

---

[8] Tyler, p. 8.

[9] Stufflebeam, p. 7.

Recommendations

1. THE UNITED STATES OFFICE OF EDUCATION SHOULD FUND A PROJECT DESIGNED TO SIMULATE THE EFFECT OF USE OF THE STUFFLEBEAM (CIPP) MODEL AS THE FRAMEWORK FOR THE EVALUATION DESIGNS OF ALL APPROVED TITLE III PROPOSALS.

If such a simulated test generates positive and practical results, a CIPP computer program should be written and used by Title III evaluators throughout the nation. Federal distribution of the CIPP model and computer program should then be made to state education departments, Title IV regional laboratories, and Title III regional centers as a recommended standard for Title III evaluation designs.

2. THE UNITED STATES OFFICE OF EDUCATION SHOULD FUND THE WRITING OF A PROGRAMMED LEARNING UNIT DESIGNED TO TRAIN TEACHERS IN A VARIETY OF RESEARCH MONITORING AND MEASURING SKILLS.

Such action-oriented researchers as Stake, Guba and Stufflebeam could be commissioned to produce such a program with follow-up inservice training available through Title III and Title IV service centers.

3. THE UNITED STATES OFFICE OF EDUCATION SHOULD FUND THE PRODUCTION OF A SERIES OF VIDEOTAPES AND FILMS DESIGNED TO TEACH ESSENTIAL RESEARCH

SKILLS TO TEACHERS AND ADMINISTRATORS.

These materials should be distributed free of charge to all state education departments, university film and television libraries, Title III centers and Title IV laboratories for free loan to present and future PACE agencies.

INSTITUTIONALIZATION OF EVALUATION[*/]

We live in an age of analysis. We also tend to synthesize and systematize everything we analyze in order to solve problems. These processes of analysis, synthesis and systemization are some of the power tools of our high-order scientific-technological society, aided, of course, by such things as computers and punched cards which supply data about our bill-paying habits, our blood types and our penchants for blonds, brunettes or redheads.

At the moment, the concern of the study group to which this paper is addressed is evaluation; not mere or abstract evaluation, but evaluation of the successes, failures, feasibilities and non- (or un-) feasibilities of various Title III innovative projects set up in school systems throughout the United States. Funds for this effort are supplied by the United States Congress and the projects are administered by the U. S. Office of Education, monitored by the several State Departments of Education, and worried about by local school administrators.

This interest in evaluation, in my opinion, must be seen in a technical-social-political context within the entire educational enterprise. The age of analysis in which we live is generating an age of assessment in education. Thus, we have a campaign developing for

---

[*/] Prepared by James D. Finn, chairman, Department of Instructional Technology, and professor of education, University of Southern California.

a national assessment program (how well are the schools doing?); several states are also asking and answering the same question within their borders.[1] Other considerations (and pressures) aside, it would be no surprise, therefore, to see this evaluation zeitgeist penetrating the Title III program. How well are all of these Title III projects doing?

Further, as technological patterns of thinking and processing invade the previously primitive (from a technological point of view) educational culture, it is inevitable that a drive for systemization should begin. For example, Hammond opens his paper on evaluation with the statement, "The need for a systematic approach to the evaluation of innovations has become one of education's most pressing problems."[2] Analysis of a sophisticated variety must precede systemization, and such analyses can be found not only in Hammond but in Clark and Guba, Guba, and Stufflebeam,[3] to refer to some very recent examples. It should be noted that all of the work cited is exceptionally rigorous and highly sophisticated. It provides an excellent base from which to attack certain practical evaluation problems, not only for Title III projects but for any instructional process; this work is in the high technical tradition and, as such, is relatively new to professional education.[4]

---

[1] There is the little matter of the very embarrassing performance of Los Angeles school children on reading tests, for instance.

[2] Robert L. Hammond. Evaluation at the Local Level. Tucson, Arizona: Project EPIC (mimeo), undated, p. 1.

[3] David L. Clark and Egon G. Guba. An Examination of Potential Change Roles in Education. Bloomington, Indiana: The National Institute for the Study of Educational Change (mimeo), undated.

Egon G. Guba. The Basis for Educational Improvement. Bloomington, Indiana: The National Institute for the Study of Educational Change (mimeo), July, 1967.

Daniel L. Stufflebeam. The Use and Abuse of Evaluation in Title III. Columbus, Ohio: The Ohio State University Evaluation Center (mimeo), July, 1967.

Further, another sign of the analysis-synthesis-system approach to evaluation in education is the continual invention, development and refinement of instruments for use in evaluation processes. Test-makers are everywhere, inventing measuring devices ranging from pencil and paper tests to simulators.[5] Guba suggests many new measuring and feedback instruments are needed.[6]

The general objective, then, seems to be in the direction of systemization of evaluation procedures. Sharp analyses, increasing and better instrumentation, process studies are all leading to Dr. Hammond's "systematic approach" to evaluation in general, and, if this study committee is any indication, to a systematic approach to evaluation for Title III projects. Such a movement to system should lead to a great deal of improvement.

---

[4] It may be a sign of age, but the writer can remember when the word "evaluation" was used as an excuse in parts of the educational community to avoid rigorous research; evaluation meant that anything went—and, in many places today, it still does. The new tradition will obviously change things for the better on this point.

[5] This phenomena can be seen among technologically oriented graduate students. One of my students completed a study on the evaluation of visual material by photographing and then measuring the eye pupil size of the evaluator and comparing it with his stated evaluation. Another is going to measure pulse pressure in much the same way. In both cases, the instrumentation had to be developed. Both of these projects originated with the students themselves. It is no accident that Egon Guba (cited above) did some very complex studies of television using an eye-movement television-film setup a few years ago.

[6] Guba, op. cit.

I would like to point out, however, that this movement toward system in evaluation also may not lead to improvement. For, in order for an evaluation system to be applied across the country, it is necessary first to institutionalize it; this is to say that, unless other means are invented, the evaluation system must be initiated, monitored and controlled by a bureaucratic system. Institutionalization of the evaluation process could destroy the innovative possibilities of Title III.

There is nothing inevitable in this potential destruction of the innovative process by the technical organization[7] for evaluation. However, if the evaluation processes as institutionalized are not to be made into a missile system aimed at the heart of educational innovation, additional analysis and invention is absolutely necessary. The remainder of this paper will examine this problem and, in addition, report some observations on the evaluation provisions of a number of Title III proposals which were studied in some detail; hopefully, the problem examination and the proposal examination can be tied together to develop some recommendations to close the paper.

---

[7] I should make explicit that I believe the evaluation process discussed above is a sub-technology within the broader concept of instructional technology.

Why Evaluation Anyway?

The basic question that needs to be asked to begin this analysis is: why is evaluation important in the educational enterprise? There are at least five purposes or reasons that can be presented in answer to this question. These reasons are: (1) to add to the substantive knowledge of educational processes, (2) to provide information in order to adjust, discard, or otherwise change the application of an on-going educational process, (3) to provide justification for a political-social-economic action relating to education, (4) to create a product (usually paper) which can move through educational bureaucratic systems and thus keep these systems operative, (5) to provide instruments which may be used to carry information on the success of the process to the educational community. These five purposes do not necessarily operate in a discrete fashion; in other words, in any one situation several may appear in the form of a mix. It is fairly easy, however, to identify the emphasis in each case. The five purposes will be briefly discussed in the paragraphs that follow.

The distinction between the first and second has been noted by many of the recent analysts, such as Stufflebeam and Guba. The first, measurements conducted under carefully controlled conditions, theoretically provides material for the corpus of educational research; as such, the results should add to the substantive understanding of educational processes. And, as has been pointed out many times, such results are rarely directly applicable to the problems of the practitioner and are of little use to decision makers. It is possible, however, for such product oriented research to come into existence as a fall-out or byproduct of a much more comprehensive evaluation procedure. The use of such research techniques as the only means of approaching evaluation has been amply criticized in recent years.

The second purpose for evaluation is now thought to be the most important when examining on-going innovative projects in education, such as those set up under Title III. Here the decision-maker gets information on a feedback system which tells him how well the process is going, what changes need to be made, etc. Various models of this evaluation procedure have been proposed (see footnote 3).

This feedback evaluation system (if it may be called that), designed to aid decision-makers dealing with practical educational problems, such as the operation of a Title III innovative center, has not yet been criticized to any extent due to its novelty for the field of education and the careful construction of the emerging theoretical models.

However, the feedback evaluation system is open to criticism. It assumes, at the outset, that the decision making process in a given school, school system or other educational entity is rational.

It is not. The folklore of education is filled with examples of the school business manager selecting curricular materials, the high school dean of women throwing Salinger out of the library, and others too numerous to mention. Prior questions have to be asked whenever the feedback evaluation system is proposed. Who or what group is the decision-maker? How does the power structure really work? What are the motivations? Unless these questions are answered and the rationality or irrationality of the particular system is analyzed, the beautiful, precise and rational models of the feedback evaluation system will not work—or, at the very least, work very imperfectly.

The third purpose for evaluation is the purpose of justification. In this case, a board of education, a state legislature, a committee of Congress or numerous other bodies both public and private need information in order to take some action respecting education. This action may be in appropriating funds, hiring additional remedial reading teachers, purchasing a language laboratory, etc. Or, in the opposite case, it may be to fire the superintendent, set the building program back two years or reduce the audiovisual appropriation by forty percent. These actions are justified by evaluation, whether formal or informal. In the case of the disposal of school personnel, the evaluation before action may be choleric and personal; increasingly, however, as statistics become everyday playthings of the mass media, justifications for political-economic-social educational action are couched in scientific garb, whether really scientific or not. We are all familiar with arguments which press in opposite directions for action by some public body based on the same evaluative report.

While it is obvious that public bodies with the appropriate authority over education have every right to evaluative information and, in fact, often need more than they get, it is equally true that the development of justification ought to be a secondary objective of educational evaluations carried out by professionals. This, of course, is a value judgment. It can be argued, for example, that the effort expended in developing a particular kind of justification evaluation in order to save a program known to be good is more important in the real world of politics than more technically adequate professional evaluation. The answer to this problem, it seems to me, is to pay attention to the two elements in the old cliche about the tail and the dog. An evaluation program set up only for justification purposes is unprofessional; a professional evaluation procedure on a program that is demonstrably good ought to develop sufficient data to justify its continuance or expansion.

The fourth purpose for evaluation recognizes the reality of the new industrial state—the corporate society. Such a state produces hierarchical bureaucracies (this phrase is, I suppose, redundant) in industry, government, labor unions and volunteer organizations as well as universities. Evaluative reports are, of course, necessary for the proper functioning of the enterprises which are the concern of these bureaucracies, particularly for the use of the technostructure, as Galbraith has called the decision-making groups in large industries.

The reader is reminded, however, that bureaucracies lead a life of their own that is somehow magically related to the flow of paper in and out of little wooded or wire baskets and conferences in conference rooms concerning the leapfrogging of this paper among the baskets. Paper, then, must be generated so that the system may lead its organic, inward life. Evaluation studies may be a large part of this pulsing circulatory system—the corpuscles, so to speak. It is emphasized that the relation of this particular form of corpuscular paper with the real, operational world may be nil or almost nil. In many cases that is not its purpose of existence.

It then follows that a careful distinction must be made between required evaluation which is necessary and has an effect on operations and decisions and that which only serves the life function of the bureaucracy itself. The first needs improvement; the second needs to disappear.

Finally, the evaluation process is undertaken to provide data on new developments in order that these data may be diffused throughout the educational community so that schools in distant places may understand and take advantage of the findings. This idea is a little tricky, as it could be held that the evaluation comes first and diffusion follows as a matter of course. In many cases this is, in fact, what happens. However, there are other cases in which the distinctions between evaluation process and diffusion instrument are not so clear. The generators of a good idea want to sell it. The evaluation can be the package. Obviously, such a package may not be the same as an evaluation package for researchers, decision-makers or bureaucrats.

These broad purposes for evaluation do not coincide very well, I fear, with the meticulously drawn detail in the charts of the experts. It may be that their only value is in delineating the perceptions of the responsible administrator on the firing line. Thus, for example, if a request for evaluation is seen as necessary for the functioning of bureaucratic life, it will be developed with that purpose in mind—a useful paper corpuscle designed for the bureaucratic arteries, not for real time operations.

It may be more important, however, to examine the drive toward systematic evaluation in the Title III program from the point of view of these five broad categories of evaluation purposes. Evaluation of Title III for what? For diffusion, for checking and adjusting ongoing processes, for substantive knowledge, for justification—perhaps of the entire Title III program itself—or for improving the circulation of a bureaucracy? I believe that the mix of these purposes must be carefully measured before an intelligent judgment can be made concerning any agreed-upon evaluation procedure.

The Problem of Institutionalization

Leaving the question of purpose open for the moment, we can turn to what I believe is the heart of the matter, namely, the question of institutionalizing the entire Title III evaluation process.

If this is so, then the arguments introduced in the introductory material could stand further examination. It was stated that systematic evaluation was considered desirable; that this was part of the general drive toward analysis, synthesis and systemization within the educational culture; and that such systemization had its bad aspects as well as its good aspects.

Although I have nowhere seen the concept verbalized in a precise manner, it seems clear that we are being asked to provide guidelines for the institutionalization of evaluation for Title III projects. The reasoning seems to go something like this: (1) present evaluation procedures are not good; they are spotty, at times sloppy and unscientific; many times they imitate the researcher's controlled experiement when they should be providing the decision-maker with feedback information to correct the system, and a credibility gap exists on the diffusion front; (2) "hard data" must be developed for public bodies at all levels concerned with the Title III program; (3) the rather bad evaluation procedures now in use are neither generating hard data nor helping in decision-making due to lack of knowledge, guidelines and skill among the operators of Title III projects; (4) by a thorough tightening up on the evaluation guidelines to be developed by experts, and by institutionalizing these guidelines with a system of information and controls, the evaluation procedure will be helped, public bodies will be made happy by the presence of hard data and better decisions will be made in directing ongoing projects. This is, indeed, an enticing picture, and to raise questions about its fundamental premises seems to be akin to questioning the institution of motherhood, Sigmund Freud to the contrary.

However, I would like to question the entire concept of institutionalization of the Title III evaluation process and insert into the record a few arguments that might at least suggest institutionalization in a different form. It is granted at the outset that, with over 1,200 Title III projects on which a considerable sum of money is being spent and with the great need to develop viable educational innovations which can be adopted by the educational community, improvement in evaluation procedures is a necessity. Further, operations on the scale of Title III require quality controls which are only made possible by large scale systematic evaluation procedures.

Granting all this questions may still be raised and arguments considered. First, while the analyses of the experts—Guba, Stufflebeam, Stake, et al—of the evaluation process are impressive and potentially fruitful, is it possible that they have, in fact, over-analyzed the process and, in doing so, slipped into the same trap that the conventional educational research man does when he attempts to apply controlled research techniques to evaluation processes operating under field conditions? Are these analyses rather important additions to our substantive knowledge and should, they instead, be used to generate more study of the process so that field applications would eventually develop? Have, in fact, these analyses departed from operational reality, at least in the sense that the practitioner would not know what to do with them? And, if one or more of these models was frozen into enforced guidelines, would this not result only in bureaucratic paper? I am not sure as to the answers to these questions, but I feel that these possibilities deserve more consideration than they have been getting.

I have no question, however, on another point. The proposed models simply do not embrace all Title III projects. There is a tendency to forget that a portion of the Title III effort is designed to provide supplementary educational services to various geographic areas, and proposals have been submitted and projects funded for such service centers. Further, entire new educational program efforts ("A Six County Program in the Performing Arts") do not lend themselves too well at first to measures that are meaningful and always present difficult problems for evaluative information systems.

The service centers present the real challenge, however. There has been, I believe, a tendency on the part of the U.S.O.E.

to play down, if not ignore, the meaning and importance of the service center concept to units of the educational community. This is due, no doubt, to the decision which placed the emphasis on the innovative aspects of Title III rather than on the supplementary service center idea. It is still a fact, as noted above, that some of these centers have been approved and funded, and if there is to be systematic evaluation for Title III, these projects must be included.

In this connection, there are two problems. The first problem is somewhat technical. The difficulty lies in the fact that the sophisticated evaluation models do not exactly fit the problem of evaluating a service center—for example, a media center supplying media services to a group of school districts. These models, for all their claim to generality, tend to concentrate on innovation in the instructional process—curriculum, methodology, the mediation of instruction. It seems pretty obvious that when you set up a service center of some kind, the distance between the regional center and the student is highly attenuated from the point of view of evaluation—both as to time and distance. To expect the evaluation process as abstracted in these models to cut through from a regional center to a student in the fifth grade of George Washington school in one of six school districts, define the effect that the sudden adquisition of a film library had upon him in a year's time, and adjust the content or service of the library accordingly is also to expect that the films will be delivered to the school via flying saucers piloted by little green men.

I should hasten to add that the principles inherent in the models can be, in many cases, applied to the evaluation of service center operation. The problem is that, if the institutionalization of the evaluation process continues to proceed and harden along the lines it is apparently proceeding, harrassed administrators will be asked to evaluate a service center operation by standards that ought to be applied to the evaluation of a new approach to phonics in the teaching of reading. This simply would not make sense.

In addition, the media world is not without a certain sophistication in the evaluation of service center operations—and these procedures relate to the principles enunciated by Guba, as one model maker, but not to the tactics that seem to be implied. To cite a homely example, if film keeps coming back into the library from a given district all chewed up, the center director then has a practical measure readily at hand which requires further investigation immediately. He must find the answers to such questions as: How are the films projected? By students, teachers or both? If teachers, are they doing it properly or do they need some training? If the human factor is not the problem, what about maintenance of the equipment? If the provisions for maintenance and control are all right, what about the performance of the Acme Repair Company on the projector service contract? Etc. Etc. Once these questions are answered, then changes can be made. This procedure is in line with the principles of Guba, Stufflebeam, et al, but not the suggested tactics that seem to flow from them. Problems such as this one will only become difficult if evaluation is institutionalized. Under such hardening of the categories [*] the evaluative universe is interpreted to be the instructional process and the responsible administrator is forced to proceed accordingly.

The second problem defies the models. An examination of proposals for Title III centers (mainly media service centers), both funded and unfunded, shows immediately that the funds are very badly needed to supply materials, equipment and services that are sadly lacking in the districts to be served. NDEA funds, articles about media, and fears about commercial domination to the contrary, the plain fact is that many, many schools in this country do not have enough of anything to do the job required of them. Under Title III they get some money for equipment, materials and services. It is like giving a drink of water to a man who has spent three days on the desert without it. How are you going to evaluate that? By the test of survival? What is survival in the educational setting? What these few centers funded by Title III mean is that all of the schools involved are experiencing an increase in their technological base. I submit it is only after this base has been functioning to the point where it requires additional technology does evaluation become meaningful. In the beginning, anything is better than nothing. There are, of course, still evaluation questions, generally pointed toward improving operations and attaining efficiency. Other models, however, are needed for this.

---

[*] A phrase picked up from Edgar Dale many years ago.

In a sense, the broad educational programs mentioned above fall into the same category as the service center. In some of the proposals I examined, for example, broad programs in the performing and plastic arts and the humanities were proposed for regions which had absolutely nothing of this kind but the prints sold at Woolworth's, the local piano teacher and the county pioneer pageant at the fair each fall. In one rather large area, for example, there was no school except a religious high school where a student could get instruction in the playing of any stringed instrument. Again, the desert-water analogy holds. Some things are obvious. Music, art or the theater brought into a community make things better, period. It seems, in a way, ridiculous to measure or count such efforts; members of Congress should be happy with the invasion of the arts as a happening; experimenters or journalists might have to wait a few years for experimentation or diffusion. It is granted that the best possible operation is needed, but a narrow institutionalization of the evaluation process for Title III projects will not provide that better operation.

If the evaluation theory we are apparently following does not exactly fit the service center and large program projects, the incongruence must show somewhere. It does in reference to objectives, an important point to notice when thinking about evaluation. Almost every reference to objectives (the achievement of which are to be measured) refers to behavioral objectives or some variant thereof. In addition, performance tests, criterion tests, etc. are easily picked up bywords in discourse on evaluation. It is as if the jargon of programmed instruction has suddenly become the lingua franca of all educational evaluation—or, for that matter, of all education.

Now, some of my best friends are behavioral objectives, but I would not want my media service center to marry one. Seriously, objectives are one thing and behavioral objectives are another. Behavioral objectives are a microcosm, to be entered into when students are directly related to content, processes, media or people in the classroom. To apply them to large programs embracing all of the arts throughout a wide region, a library service center for a county school system, or a data processing installation represents a beautiful confusion of form with substance—setting up the conditions of operation for an educational Parkinson's Law. Even smaller sectors of the educational enterprise directly related to instruction may not need objectives stated in behavioral terms.

I wish to make it very clear that I am not attacking behavioral objectives as such. They can be made to accomplish spectacular things with certain instructional processes and are legitimate targets for evaluation. On the other hand, sometimes we will need system objectives, which are not the same thing at all. By stretching a point, it might be said that an evaluator might want to measure (and change) the behavior of a system (such as a library, a full-scale curriculum operation or something else), but I do not believe that this type of objective was exactly what B. F. Skinner had in mind (or Ralph Tyler many years earlier).

There is, of course, nothing in the models with which we have been dealing that requires behavioral objectives; and it is also a truism that evaluators can't evaluate for any purpose without objectives. The fact remains that all the discourse about evaluation is conducted as if there were no other types of objectives in the educational universe, even when the discussants unconsciously know better. Thus the incongruence between the theory (or theories) of evaluation under analysis and the real time world of operation can be shown to be a possibility.

This exploratory discussion relating to some specific problems of institutionalizing Title III evaluative processes can now be brought into focus at the philosophical level. To review, there is apparently great concern as to the quality of existing evaluation, there is a desire to produce "hard data" for persuasive purposes, there is a need for accurate information as to progress and to adjust for improvement, and there is the necessity of diffusing information on successful practices—success being determined by competent evaluation. Further, the size of the Title III effort (over 1, 200 units) and its wide distribution geographically with enormous differences in the resources and abilities of the educational units involved, all press for standardization (at an acceptable level of competence) of evaluation procedures.

There is, however, a deeper drive involved in this effort—or, at least, I believe it to be so. The industrial state is the corporate, bureaucratic state. The imperatives of technology, we are reminded by many observers such as Galbraith, have replaced ideology in much of our culture. Technology requires large scale organization, orderly processes, group planning, and, where possible, it seems to me, a kind of neatness in the system that one might associate with a computer installation or a "clean room" in an electronics factory.

In a fundamental sense, the entire Title III effort at educational innovation is a move toward bringing the educational enterprise into the modern industrial state. Title III _is_ educational technology. This may sound strange to those educators who define educational technology as a term synonymous with language laboratories, computers or television. I would remind them of Galbraith's definition of technology, although many similar definitions might be cited. Galbraith said, "Technology means the systematic application of scientific or other organized knowledge to practical tasks."[8/] He goes on to point out that the main characteristic of technology is the breaking down of tasks into detailed sub-divisions so that organized knowledge may be put to work, and that

---

[8/] John Kenneth Galbraith. _The New Industrial State_. Boston: Houghton Mifflin Company, 1967, p. 12.

this analytical procedure "is not confined to, nor has it any special relevance to mechanical processes."[9] I submit that the very selection of evaluation as a field to analyze in connection with the Title III program is evidence of this movement toward technology.

In any of the units of a society of high technology, such as the United States, very extensive planning is necessary. Since the units are very large, for the most part (Galbraith's 500 "mature corporations," large government, etc.), the planning affects and controls millions. A systemization of the Title III evaluation process is a form of planning. This concept leads us to the crux of the argument. Galbraith has noted, for example, that ". . . planning involves, inevitably, the control of human behavior. The denial that we do any planning has helped to conceal the fact of such control even from those who are controlled."[10] He was speaking of economic planning, but I believe this concept to be totally generalizable in our technological culture.

Planning involves the creation and management of systems; systems require, or at least imply, bureaucratic control. Hence, unless, as indicated in the earlier portion of this paper, additional means are invented to fit the peculiarities of the institution with which we are dealing—the American educational enterprise—the development of a systematic, technically competent evaluation process for Title III will result in bureaucratic control that I believe would mean the end of the dream that Title III would bring needed innovation to American education.

---

[9] Ibid., p. 13.

[10] Galbraith, op. cit., p. 23.

Such a prospect is difficult enough, but further complexities must be examined. Galbraith has pointed out that, with high technology and large organization, as in the mature industry, the planning and management processes are in the hands of fluid groups of experts, each bringing complex information into the group processes where decisions are made. The fluid groups he calls the technostructure. We come now to the rub. American education, as a sector of the political economy, is very primitive from a technological point of view and has practically no techno-structure.

If the concept of a lack of a technostructure in American education is accepted, it is possible to explain many things.[11] Galbraith does this in another context when explaining why socialist countries have had "the most uniformly dismal experiment of countries seeking economic development."[12] Speaking of India and Ceylon, he goes on to say that, in these countries,

> ...if the minister is to be questioned, he must have knowledge. He cannot plead that he is uninformed without admitting to being a nonentity.... Technical personnel are

---

[11] Consider the inability of the old line-staff administrative patterns to handle aspects of the new educational technology (hardware and materials logistics, etc.); consider the problems of the ghetto from this point of view; etc.

[12] Galbraith, op. cit., p. 101.

> less experienced than in the older countries. Organization is less mature. These lead to error, and suggest to parlimentarians and civil servants the need for careful review of decisions by higher and presumably more competent authority. Poverty. . . calls for further review. And rigid personnel and civil service rules, the established British answer to primitive administrative capacity, extend into the public firm and prevent the easy constitution and reconstitution of groups with information relevant to changing problems.[13]

It seems to me that it is easy enough to transfer this Galbraithian concept to the American educational system. To begin with, there is no large scale organization in the technological sense and, as noted, no technostructure. If development is to occur, it becomes obvious to those responsible—in our case, the U.S.O.E. and our study committee—that review and control of decisions and operations relating to evaluation are absolutely necessary when dealing with such "primitive administrative capacity"—and it _is_ primitive from this point of view.

At this point, however, it is necessary to exorcize a ghost. I am suggesting that national bureaucratic control or even systemization of evaluation seems to be necessary under the circumstances, but I am further suggesting that this may be unwise (the reasons for this will be discussed below). It then might follow that all I am interested in is a reduction in the size of the bureaucracy and the removal of the controls to the state level. Nothing could be further from the truth.

---

[13] Galbraith, _op_. _cit_., pp. 101-102.

Bureaucratic control from the state level will merely extend the "primitive administrative capacity" from the school district upward. I venture to say that no state department in the United States has an adequate technostructure or is about to get one—and this includes New York and California, both of which have been praised in many quarters. I believe that institutionalizing Title III evaluation processes under state departments of education will concentrate many undesirable elements of such a system. My arguments against institutionalization in the form that seems to be implied by events must be seen in this light. Once the arguments are considered, it may be possible to suggest a better solution.

To return to the main theme, given the assumption of planning and systematic (translate bureaucratic) control of Title III evaluation procedures, certain undesirable effects seem to inevitably flow from many (but not all) such developments. High technical solutions to some problems require exactly such arrangements—getting to the moon, stamping out an epidemic, etc. In such situations large scale technology and its peculiar requirements seem to fit fairly well and the people involved are relatively comfortable. However, all large scale applications of technology (systematic organization) do not fit—particularly where they impinge in certain ways on human beings (recall Galbraith's sentence on control).

This lack of human fit of many of the technological developments in the United States—depersonalization of university life, smog and the automobile, social decisions made by the corporate structure over which those being decided about have no control—has given rise, in the last decade, to a heated dialogue which has erupted on numerous occasions into violence.

It is a mistake, however—a serious mistake—to view the dialogue only from the point of view of the violence or certain individual issues such as Viet Nam, civil rights or rent strikes. For anyone who cares to take the time to inquire, a much deeper dialogue, a much deeper emerging philosophical statement are there to hear. "To hear" is used advisedly, for I believe unless we listen to what some of these bright young people are saying, to what the New Left is trying to expound, to what some artists are expressing, we, as educators, may fail this country and all of the young people in it.

What is this dialogue? It is a dialogue between high-order technological organization, the industrial state, impersonal controls over people and spokesmen, no matter to the degree that they are right or wrong, for men as human beings—for man in microcosm. An educator, it seems to me, does not necessarily have to adopt totally the view of one side or the other. Some educators, at least, ought to see the thousand dilemmas present in this confrontation and seek solutions which are, first, educative, and secondly, human without reducing our culture back to some primitive stage where we live in the hills in shacks. I believe we should seek in general what William Javanovich saw in the future when he predicted "the emergence of a new kind of intellectualism which will reconcile content with style, social purpose with personal sensibility."[14]

Assuming that I am right in understanding that the effort to "improve" the evaluation processes associated with Title III will move in the direction of national systemization and control (or, worse, state systemization and control), the criticisms of these spokesmen for the defense of man as man have relevance. They should be seriously thought about, for, within the intimate environment in which each man lives, they attack systemization and control with a vengeance.

Let us begin with one of the best known spokesmen for this point of view, Paul Goodman. Recently he was asked to address the National Security Industrial Association and took the opportunity to berate this industrial-military technology group. At one point he said:

> Your thinking is never to simplify and retrench, but always to devise new equipment to alleviate the mess that you have helped to make with your previous equipment.

And, then he went on:

> Your systems analyses of social problems always tend toward standardization, centralization, and bureaucratic control, *although these are not necessary in the method*. (italics mine)

Finally, he stated a principle or theme that reappears time and time again in this literature:

---

[14] William Javanovich. "My Illusions and Yours." *Harpers*, Volume 235, Number 1409 (October, 1967), p. 59.

> In a society that is cluttered, overcentralized, and overadministered, we should aim at simplification, decentralization, and decontrol.[15]

A great deal of this new literature is being created by young people. A whole issue of the American Scholar was recently devoted to writing by people under thirty. In it, Michael Rossman made an effort to explain the deep philosophical base of the so-called National Student Movement. In doing this, he expressed much about their concern with man as individual man and even explained (and this is a little hard for an older person to understand) "participatory democracy" both as philosophy and as tactic. Three concepts appear in much of this literature, and they appear in Rossman. They are: Engagement, Encounter, and Involvement. The concern is with humans relating to humans—with true encounter. Rossman puts it this way:

> . . . the present Old Left among us... aims at the mass; at the racial, economic or occupational population. But the unit in terms of which the Movement conceives change tends to be the small group.
>
> ...The way to influence large groups is by local example, rather than global persuasion.
>
> ...direct personal involvement is the Movement's human backbone.
>
> ...In saying that people must be involved in the decisions that shape their lives, the emphasis in on involved.

---

[15] Paul Goodman. "A Causerie at the Military-Industrial." New York Review of Books, Volume 9, Number 9 (November 23, 1967), pp. 16-17.

> ...political dialogue must be cast in a different vocabullary than that possible with the comfortable separation of the Changer and the Changed.[16, 17]

Harper's recently published a symposium consisting of a series of dialogues between well-known older commentators on the national scene and a panel of young people similar to those appearing in the American Scholar. In this case, it was, at times, hard to distinguish between the older and younger viewpoints. Again, the theme reappeared time and again. The issue was impersonalism, bureaucratic control *versus* general encounter and the human condition. It is tempting to go on quoting a great deal because the material seems so relevant, but I shall try to restrain myself. Paul Potter, one of the "older" members of the panel, said:

> ...there is a growing belief that the only force really shaping the future is the force of unleashed technology controlled by giant, impersonal bureaucracies.
> ...economic planners cluck truculently about the "great leveling force of technological development" that will in time assimilate all revolutions and all cultural diversities into one grand machine civilization.[18]

In commenting on Potter's article, young Robert Gross said:

> ...we have to end the domination of this society by the

---

[16] Michael Rossman. "The Movement and Educational Reform." The American Scholar, Volume 36, Number 4 (Autumn, 1967), pp. 595-596.

[17] It should be noted that, for the professional educator, Rossman has some provocative things to say about teaching and learning in higher institutions and proposes some interesting reforms.

[18] Paul Potter. "The Future is Not Inevitable." Harper's, Volume 235, Number 1409 (October, 1967), p. 48.

> large, rigid bureaucracies which pay little attention to the needs of the people they are intended to serve.[19]

And Alfred Kazin commented:

> ...the more immediate and abundant our technical power, the more we lose the naive, spontaneous imagination.[20]

Other parts of the text refer to "students who are demanding flexibility and personal relevance," "non-rational ways of getting at knowledge," "taking strength from the free private life." There is also, however, another thread which suggests that something better might be made of this "technocratic totalitarianism," as Potter called it, and he went on to say:

> The technology and the bureaucracy can be mastered and put to work to create for everyone what we've begun to have a taste of.....[21]

A Suggested Accommodation

The argument has come full circle, and the potentials of accommodation are there if they can be identified. It seems to me important to suggest a new approach to the problem of institutionalizing evaluation for Title III projects which would accommodate need for and technique of consistent, high-quality evaluation procedures with human, local needs and differences in projects and concepts. Personally, I feel that many of the critics of technical bureaucratic control cited above offer little as replacement for this control with a sort of leaderless "participatory democracy" which, in a technical sense, will not even achieve their own objectives. And yet, much of what they have to say is important.

---

[19] Robert A. Gross. "To Mr. Potter." Harper's, Volume 235, Number 1409 (October, 1967), p. 50.

[20] Alfred Kazin. "Art on Trial." Harper's, Volume 235, Number 1409 (October, 1967), p. 51.

[21] Potter, op. cit., p. 50.

We have been concerned with models of high-quality evaluation procedures; with purposes of evaluation; with implied arrangements to insert controls in the system so that legitimate purposes may invariably be supported with technical competence. All of the elements of a bureaucratic system are there—whether in the eventual rough and tumble of administrative or legislative politics this control is placed at the Federal or state level. Of course, such control could apparently be non-enforced by guidelines or some other system which in fact would quickly encrust into a strait-jacket. On the other hand, controls are needed so that competent, useful evaluation may take place. This, to repeat, is the problem of accommodation.

And I hold that it is a solvable problem and that the possible solution, as Boyd Bode used to be fond of saying, "lies at hand." Many of the elements are present in Project EPIC of Tucson, Arizona.[22/] Project EPIC is, essentially, a sort of local evaluation service center funded by Title III funds and assisting local school agencies within the area it services.

[22/] See Hammond, op. cit.

Given this idea as a start, it is possible to make a series of recommendations that can, I believe, achieve the sought after accommodation between the need for evaluation and the human variation which inevitably occurs at the end of the line. Such an accommodation will not be as neat as a clean room in an electronics factory; on the other hand, it will not be so messy as to be useless; in fact, it might have enough variation in it to release creative energy—which was the general idea of Title III in the first place.

Recommendations

1. TITLE III FUNDS BE USED TO SET UP A SERIES OF REGIONAL EVALUATION CENTERS THROUGHOUT THE UNITED STATES DESIGNED TO PROVIDE TRAINING AND ASSISTANCE TO LOCAL EDUCATIONAL AGENCIES.
2. THE FUNCTION OF THESE CENTERS BE TO PROVIDE ADVICE, TRAINING AND SERVICES AND, PARTICULARLY, TO DIFFUSE THE GENERAL IDEA OF THE IMPORTANCE, USEFULNESS AND NATURE OF A HIGH-QUALITY EVALUATION SYSTEM.
3. IT BE UNDERSTOOD THAT THE EVALUATION CENTERS ARE ONLY PERSUASIVE AND HELPFUL IN NATURE AND THAT, IF AN EDUCATIONAL AGENCY CHOOSES NOT TO

RESPOND, IT BE ALLOWED TO WITHOUT PENALTY— ACTUAL OR IMPLIED.

4. THESE CENTERS ALSO ENGAGE IN A CERTAIN AMOUNT OF APPLIED AND FIELD RESEARCH WITH THE PURPOSE OF DEVELOPING VIABLE AND VARIABLE EVALUATION PROCEDURES WHICH CAN EMBRACE ALL TYPES OF EVALUATION NEEDS AND PURPOSES.

5. A BACK-UP NATIONAL BOARD BE SET UP TO ASSIST THE CENTERS AND THE U.S.O.E. AND CONGRESS. THIS BOARD WOULD HAVE THE FOLLOWING FUNCTIONS:

   a. Locate and rotate manpower between the centers. Much of this manpower could be one-year leave-of-absence type; other slots could be filled with qualified graduate students on an intern basis.

   b. Act as the assembling agency for results which ought to be diffused and as the communication agency between the centers. As such it should act as both the stimulus and the conscience for the centers.

   c. Engage in broad scope research and development studies in the field of evaluation.

   d. Provide an information source for all government agencies, local, state, federal.

   e. Relate to and diffuse information to the educational community about other national, private evaluative efforts, such as the National Assessment Program, etc.

Under no circumstances should this board be thought of as a control mechanism in the bureaucratic sense.

If these recommendations are analyzed for the purpose for which they were made—to create a system which would achieve the objectives of necessary high-quality evaluation procedures for local, human purposes without inserting another bureaucracy into the system, details of operation and administration should become reasonably clear. The human being at the end of the line—administrator, teacher or media specialist—can have his opportunity for involvement and encounter. And it is highly likely that we can raise the quality of evaluation immensely.

AN EVALUATION MODEL[*]

The purpose of evaluation, Dan Stufflebeam[1] has told us, is decision making. There is no justification for evaluation where there are no decisions to be made—and where there are decisions to be made there is need for evaluation. There are many decisions to be made about the implementation of Title III, many different kinds of decisions. It is reasonable to expect that there will be many kinds of evaluation to guide those decisions.

We can represent three major kinds of decisions and evaluation needs by three questions:

What is the destination?
Which path shall be taken?
What pace shall be set?

This is a homely way of saying that we make decisions about our goals our projects, and our tactics. Many decisions about each. If our decisions are to be rational, overt, deliberate—as opposed to intuitive, covert, and impulsive—we need information. Formal evaluation is a major source of that information. As different information is needed, different evaluation plans must be available.

---

[*] Prepared by Robert E. Stake, associate director, Center for Instructional Research and Curriculum Evaluation, University of Illinois.

[1] Stufflebeam, Daniel L. "The Use and Abuse of Evaluation in Title III." An address delivered at the National Seminar on Innovation, Honolulu, Hawaii, July, 1967.

Of course, there is no one-and-only time we decide where we are going, which path to follow, or how to proceed. We make these decisions sometimes formally and with a bit of ceremony, usually informally, always repetitively, with conscious and unconscious review, checks and balances. In addition to the three questions above, then, we should add these:

> Are we headed for the destination we chose?
>
> Are we on the path we chose?
>
> Are we proceeding in the manner we planned?

These are status-checking questions, the questions of mechanical governors, the questions of self-synchronizer units in the cybernetic system. They are important questions, leading to important decision—but they are different from the first three, and require different observations and information.

Evaluation Issues

What information for what decisions is indicated in the evaluation plan. Several Grand Plans, basic models, are available. Stufflebeam presented one. Such people as Henry Walbesser[2], Michael

---

[2] American Association for the Advancement of Science, Commission on Science Education. An Evaluation Model and Its Application. Science—A Process Approach. Washington, D. C.: The Association (1515 Massachusetts Avenue, N.W.), 1965. 99 pp.

Scriven[3], and I[4] have presented others. These plans or models differ in the help they offer the decision-maker. It may be helpful here to indicate some of the issues that the different models raise. As evaluators of this colossus, Title III, we cannot be oblivious to these issues.

The desire for evaluation of educational programs is not new (nor is it a particularly strong desire among practitioners). But the current abundance of models for evaluating seems new to me. The challenge to traditional curricula since Sputnik and the increasing federal and philanthropic-foundation interest in education has stirred researchers and developers alike. John Goodlad[5] has summarized their work nicely. A concurrent concern for evaluation apparently has stirred the evaluation-model-builders. Some of their new models emphasize the use of conventional tests, others do not. Other differences are: (a) the importance of the classroom teacher as a developer of curricula, (b) reliance upon the developer's intuitive rather than rational skills, (c) emphasis on subject-matter-content goals as opposed to intellectual-process-and-skill goals, (d) whether or not assessments will be limited to the developer's stated goals, and (e) the importance of building-in plans for reassessing goals during and after the developmental phase.

---

[3] Scriven, Michael. "Methodology of Evaluation." American Educational Research Association Monograph Series on Curriculum Evaluation 1: 39-83; Chicago: Rand McNally and Co., 1967.

[4] Stake, Robert E. "The Countenance of Educational Evaluation." Teachers College Record 68: 523-40; April, 1967.

[5] Goodlad, John I. The Changing School Curriculum. New York: Fund for the Advancement of Education (477 Madison Avenue), 1966. 122 pp.

One important inventory of evaluation issues was offered by Michael Scriven. According to him the important dichotomies are:

1. Formative vs. summative evaluation, i.e., evaluation during development to check the quality of components vs. evaluation after completion to check the quality of the whole.

2. Professional vs. amateur evaluation, not only how well trained are the personnel in measurement and research methods but also are they professionally competent to deal with the subject matter, the teaching methods, the philosophical issues, etc.—and in addition, are the evaluators disinterested parties?

3. Evaluation vs. process studies, i.e., studies to discover the worth of a program against studies to discover the nature of a program, is the emphasis more on judgment or description?

4. Intrinsic vs. payoff evaluation, i.e., studies to discover the quality of the input vs. studies to discover the quality of the outcomes.

5. Comparative vs. noncomparative evaluation, i.e., studies that compare alternate programs or procedures, with or without a so-called control group, vs. those which concentrate on the experimental program or procedure alone.

6. Evaluation vs. explanatory studies—explanatory studies

are expected not only to indicate the "whys of the outcomes," as Tom Hastings[6] puts it, but also generalize beyond the specific curricula used.

Title III Evaluation Components

PACE calls for supplementary centers to facilitate educational innovation in schools throughout the country. Like any federal program[*] it has its rationale, its purposes, its participating projects, its procedures, its obstacles, and its reward system. And a network of decisions. These decisions vary in kind and in purpose. An evaluation of Title III activities requires evaluation activities that vary in kind and in purpose as well.

Let us consider the general components of program evaluation. In an oversimplified manner, four components are identified: goals, projects, tactics, and outcomes. First, goals.

Goals: The important emphasis Ralph Tyler[7] and many colleagues have given to goals will not be summarized, nor will PACE goals be identified. To begin with, goals are defined first in terms of wants and only second in terms of words. Every attempt to translate wants into words will be less than perfect. Any list of goals (statements of behavioral objectives or otherwise) is suspect either if it appears to supercede the wants or to represent them as unchanging. Any evaluation must presume that stated objectives do not perfectly represent what is wanted.

---

[*] Contrary to some official usage I am using the term "program" to indicate the entire Title III operation. The term "project" here refers to any local undertaking, simple or compound.

[6] Hastings, J. Thomas. "Curriculum Evaluation: The Whys of the Outcomes." Journal of Educational Measurement 3: 27-32; 1966.

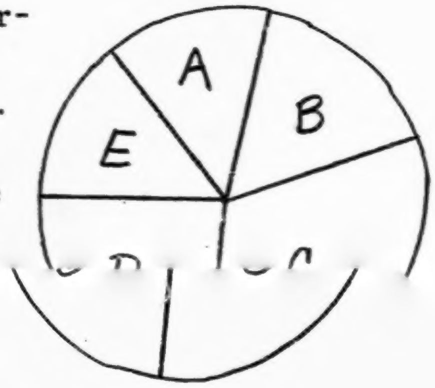
[7] Tyler, Ralph W. Basic Principles of Curriculum and Instruction. Chicago: University of Chicago Press, 1950. 83 pp.

A valid list of educational goals will contain competing and even contradictory goals. Goals are competitive in the sense that each pursuit costs something and the total of our resources will always be less than the cost of pursuing all goals. We have to choose among our goals. We assign priorities to them. We may do this consciously or unconsciously. But we do it. It's a matter of choice and we have no choice but to choose.

Goals will be contradictory. Often we seek incompatible outcomes. We try to teach faith and skepticism. We try to instill deep appreciation and yet provoke aspiration for something better. We try to give teachers opportunity to be creative, yet we try to bolster instruction through reasonable insistence on using the methods, topics, and materials of time-tested programs. We hope that any one teaching effort will aid persons with different headings. We seek to serve a pluralistic society. Contradictory goals are to be expected in a pluralistic society. We cannot hope to pursue only goals that are perfectly complimentary and universally wanted.

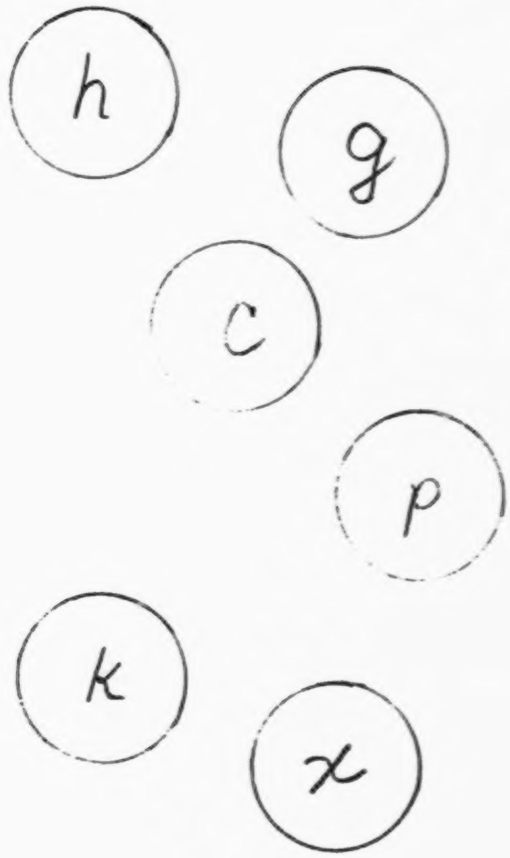
Evaluators must realize that goals are changing, competitive, and even contradictory. A program evaluation is incomplete if it ignores this fact, if it goes no further than listing the specific goals at time zero. To understand the Title III operation and to ascertain its value, we are obligated to identify goals, ascertain priorities, reveal the dynamics of changing priorities, and provide information for decisions about new goals and priorities. This is not to say that these things must happen first, before we do anything else; nor is it to say that we must be as specific as a blueprint; but as part of the operation we must obtain some communicable representation of the wants this legislation was designed to alleviate, and of the transformation of these wants into other wants over time. A representation of Title III goals is given by the figure on the right, a reminder perhaps that the Title III dollar doesn't alleviate everyone's wants.

Title III Goals



Projects: There are many pathways to the goal-destination. Title III administrators must choose among paths, among projects. More plans will be proposed than can be followed, some must be selected, some must be rejected. Each project has its goals. Some projects will concentrate on a single goal, most pursue a complex.

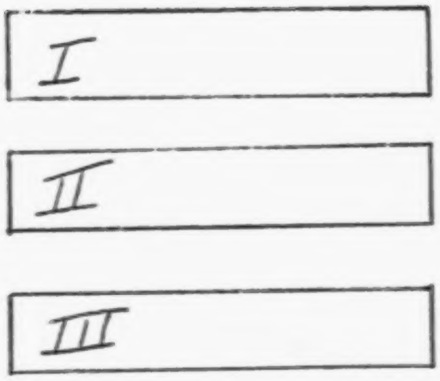
The rationale of a project may be independent in spirit—it can



Title III Projects

emphasize local problems, local resources and local remedies, but it cannot be unrelated to general goals. Each local program proceeds in some determinable ways to facilitate or obstruct the pursuit of national goals. The success of Title III is not merely the aggregate of progress and setback in the pursuit of local program goals. There is a national purpose—as represented by interpretations of the legislation—and, however, close the fit, it will not coincide with the aggregate of program purposes.

Tactics: Even when the target and path are set, there are many modes of travel. No one is necessarily right. Whether the Title III project is a multi-school dreadnaught or a lone knight-errant, each proposal cites special resources and techniques for doing its deed. One generalization we have learned from educational practice and research is that almost any technique will work, at least in some circumstances. We are often persuaded that a favorable combination of resources—community, school, and home—is more influential than the pedagogical techniques we employ. But we have no choice but to continue searching for better tactics to overcome the seemingly immutable, deleterious educational conditions in many communities. Our evaluation must consider the alternate tactics among which we search.



Project Tactics

Outcomes: Any one project can be measured in terms of its outcomes. The changes will be made: student achievement and attitudes, teacher maturation and alienation, organizational efficiency and compassion, renaissance of hope and initiative. There will be costs which, in a sense, are outcomes: expenditures of money, time and effort, drains upon public and family confidence, etc. It is certain that few changes will be easy to measure—and fewer still directly ascribable to the Title III effort. But there they are: costs and benefits, seen in different coin by different groups—various views of which are essential as description of the project and essential as an information base for decision-makers.

There is little sense in viewing the success of a project as unidimensional. There is no single criterion outcome that is paramount, nor any composite that is a satisfactory index of success. As there are multiple

Outcomes

goals there are multiple successes and failures with any endeavor. A project is unfairly considered if its record is reduced to a single value-dimension. So it is with the overall Title III program.

We seldom need such reduction. Sometimes a decision-maker must make an all or nothing decision, such as continuance or abandonment. Then some single continuum of value is an important fact of life. But most decisions pertain to degree of emphasis, here and there.

There are occasions when only one of several projects (or tactics) can be supported, where the goal priorities are similar and where each promises a different profile of outcomes or where each has registered a different profile of gains, with no profile clearly superior. When only one can be supported, a single continuum is again a necessary fact of life. The decision-maker must weight outcomes so as to arrive at a single index of desirability. Clearly, one project must be rated ahead of the others. But these are not common occasions. More often, the proposals competing for immediate funding are not so similar in goal priorities, or the superiority of one of the proposals is clear.

The Go-Round: To make a point or two, the four components are arranged together in Figure 1. The arrows indicate there is a cyclical influence, one on the other. Each component is influenced by local community conditions.

1. A rational setting of goal priorities and funding of projects rests upon knowledge of

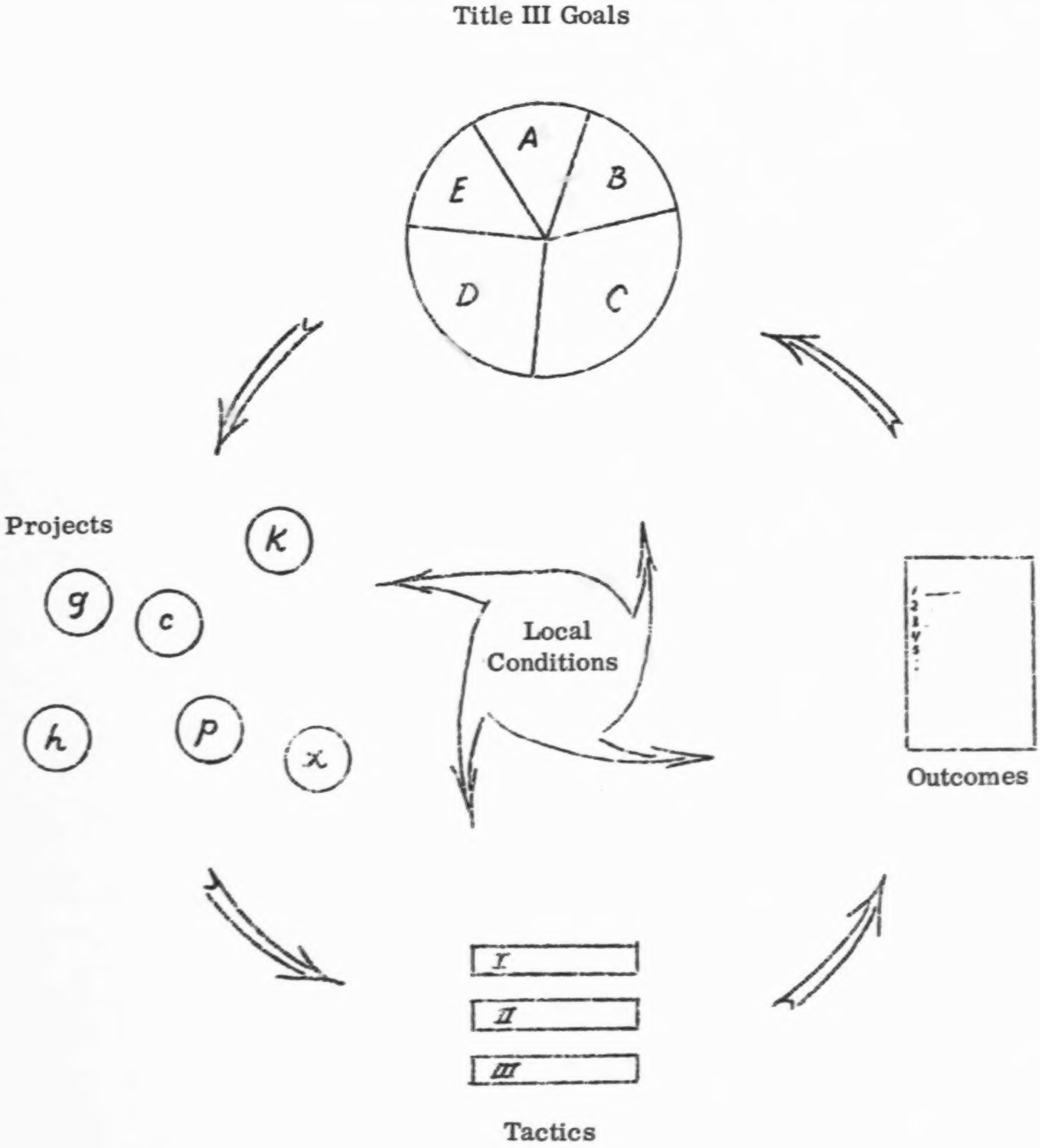

Title III Goals
A
B
C
D
E
Projects
k
g
c
h
p
x
Local Conditions
Outcomes
1
2
3
4
5
I
II
III
Tactics

Figure 1. A representation of the major components of the Title III program.

   - a. what goal alternatives (wants) there are
   - b. what resources are available
   - c. what types of projects can be funded
   - d. approximate cost/benefit ratios for various types of projects

2. Certain projects are selected for funding. By careful plan or merely in effect, a certain combination of goals is operationalized.

3. For each project different tactics are selected. The choosing of tactics continues through the duration of the project. By plan or in effect, these tactics operationalize a certain combination of project goals, and through aggregation they operationalize a certain combination of program goals.

4. A rational choice of tactics rests on knowledge of
   - a. project goals
   - b. what tactics are available
   - c. contingencies between tactics and outcomes
   - d. what resources are available
   - e. the costs of various tactics

5. Outcomes are functionally dependent on tactics, given a set of local conditions.

6. Goals can be defined as some combination of outcomes.

The principal consideration of these comments is that a sequence of decisions are there to be examined both at the federal and local level.

A Decision-Evaluation Matrix. To elaborate on the role that evaluation may play in Title III, a matrix is presented in Table 1.

Federal and local decisions/evaluations are set apart although much of the activity at those two scenes is parallel. This is not to say that the aggregate of information gathered for local decision-making will suffice for national decision-making, but it clearly is a large part of what is needed.

In addition to the six questions that were raised in the first pages of this paper, several more are raised with regard to expected and observed outcomes. Included in the observed outcomes are the "hard data" for which there is such a rowdy official appetite these days. Those outcome data are important but no more so than the other evaluation data indicated in the second and fifth columns of the matrix.

In the third and sixth columns, eight different evaluation activities are given. I have borrowed from Scriven here, and my countenance paper but I have put a new slant on some of the terms. They are given the following formal definitions:

Evaluation Activities

1. Priority Setting—A study of wants under a given rationale or philosophy leading to preferential rating of goals, with implication for implementation.

DECISION LEVEL	Initiative Decisions	Additional Information Needed	Evaluation Activities	Maintenance Decisions	Additional Information Needed	Evaluation Activities
FEDERAL PROGRAM	Which goals to pursue?	Rationale Wants National Resources	Priority Setting Environmental Survey	Are we now pursuing goals we choose?	Project goals effected	Goal congruence study
	Which projects to support?	Alternative Proposals Cost-Benefit Ratios	Intrinsic Evaluation Project Summative Evaluations	Do we have the program we thought we were getting?	Project Activities	Intrinsic Evaluation
	What results to expect?	Feasibility Local Conditions	Feasibility Study Environmental Survey	Are we getting the results we expected?	Project Outcomes	Payoff Evaluation

DECISION LEVEL	Initiative Decisions	Additional Information Needed	Evaluation Activities	Maintenance Decisions	Additional Information Needed	Evaluation Activities
LOCAL PROJECT	Which goals to pursue?	Rationale Wants Local Resources	Priority Setting Environmental Survey	Are we now pursuing goals we chose?	Project Goals	Goal congruence study
	Which tactics to employ?	Alternative Tactics Means-Ends Relationships	Intrinsic Evaluation Formative Evaluations	Are we using the tactics we planned to?	Project Activities	Intrinsic Evaluation
	What results to expect?	Local Conditions Previous Results	Environmental Survey Summative Evaluation	Are we getting the results we expected?	Project Outcomes	Payoff Evaluation

Table 1. A decision-evaluation matrix for planning Title III evaluation activities.

2. Feasibility Study—An estimation of the costs of overcoming various obstacles to implementing a given program or project.

3. Environmental Survey—A gathering of information about the setting in which the program or project will occur, including its resources, social institutions, existing programs, personnel, organization, etc.

4. Goal Congruence Study—A study of the relatedness of goals of different undertakings or of the relatedness of stated goals to those implied by practices.

5. Intrinsic Evaluation—An analysis of the logic of the plans and activities of a program or project, providing judgments of relevance and value of various components.

6. Payoff Evaluation—An empirical study of the degree to which observed outcomes approximate intended outcomes.

7. Formative Evaluation—The empirical study of the effects of various tactics, emphasizing functional relationships potentially useful to other program development.

8. Summative Evaluation—The empirical study of the effects of a whole project under given environmental conditions, preferably with comparisons to alternate projects.

My conclusion for the moment is that there is a lot more to be looked at than we have been looking at formally.

PACE EVALUATION IN GENERAL*/

This report is based upon an analysis of the evaluation procedures outlined in 21 funded Title III proposals. As such, it is suggestive rather than definitive. Seven of these proposals had been submitted for the May 1966 deadline, and hence followed the older manual which required a description of evaluation only in the case of operational proposals. Three planning proposals were included in this group, and these of course contained no explicit statement on evaluation. The remaining 14 proposals followed the new manual. The range of funding requested in the 21 proposals extends from about $21,000 to about $2 million.

The proposals were analyzed from a number of points of view. (1) Using the Stufflebeam model as a guide, they were examined to determine what, if any, provisions had been made for each of the four classes of evaluation: context, input, process, and product. (2) They were examined to find out if the evaluation procedures could meet minimal criteria of validity, reliability, timeliness, credibility, pervasiveness, and budgetary adequacy. (3) They were examined to classify the designs, the means of data collection and the populations to be sampled, the criteria that were to be invoked to interpret the data, and the agents that were involved in planning and executing the evaluation procedures.

---

*/ Prepared by Egon G. Guba, director, National Institute for the Study of Educational Change, Indiana University.

The first two of these examinations were based on the entire proposal, i.e., relevant statements found anywhere in the proposal were considered. The third examination was based exclusively on the statements made in the labelled evaluation section.

Description

1. Findings in relation to the Stufflebeam model: Despite the fact that 18 of the proposals were written in response to a format explicitly requiring some discussion of evaluation and despite the fact that the PACE Manual heavily emphasizes product evaluation, only 13 of the 21 proposals deal with product evaluation in even a minimal way. As might be expected, in most of these cases the product was viewed from the point of view of effectiveness, rather than of feasibility or quality, as Taba suggested in the report of last year.

Eighteen of the 21 proposals were judged to have made some effort at context evaluation, but this judgment was arrived at by counting as context evaluations any attempts to describe needs or problems of the region to be served by the project. Since a needs survey was mandatory, most of the proposals could be considered to have been based upon such an evaluation. In no case, however, did the proposal writers give any indication that they regarded this procedure as a form of evaluation; hence we may safely assume that they were ignorant of the fact that most modern evaluators would so label it. They seem to have done the right thing even if for the wrong reason.

Only 13 of the 21 proposals demonstrated that there had been any consideration of alternative inputs designed to overcome the identified problems and needs. Typically such consideration was very informal. Input evaluation, if it occurs thus seems to be mainly intuitive. Process evaluation was almost completely ignored, with but three of the 21 proposals giving any attention to it at all.

2. Findings in relation to criteria for judging evaluations: The proposals were examined to determine whether the proposed evaluation procedures could meet minimal criteria of validity, reliability, credibility, pervasiveness, timeliness and budgetary adequacy. With regard to the first five of these criteria, in only six cases was enough information given to permit any assessment, and in these six cases, more than half of the assessments were negative, i.e., the procedures did not meet the criteria. With regard to budgetary adequacy, only 10 of the 21 proposals contained explicit items in the budget for evaluation. Typically the item was concerned with salary provision for a person to plan and execute the evaluation or to reimburse out-of-pocket expenses for published tests. The complexities of evaluation and the costs associated with them are obviously not appreciated.

3. Findings in relation to methods: The labelled evaluation section of each proposal was examined to determine (a) the design employed, (b) the means of data collection and the populations

sampled in their collection, (c) the criteria specified in terms of which the collected data are to be interpreted, and (d) the agents charged with the responsibility for designing and carrying out the evaluation. Each will be considered in turn:

a. Designs: The matrix displayed in Table 1 was used to classify the proposals in relation to designs.

TABLE 1

CLASSIFICATION OF DESIGNS

Design Name[1]	When the Design Involved A	
	Convenience Sample	Statistical Sample
1 - Single Measurement	1a	1b
2 - Pre- and Post-test	2a	2b
3 - Control Group	3a	3b
4 - Pre- and Post-test Plus Control Group	4a	4b
5 - Quasi-experimental Design	5a	5b
6 - Other	6a	6b

No information relevant to design was given in 13 of the 21 proposals, including the three planning proposals submitted under the earlier format. Of the remaining eight proposals, five indicated Design 2a (pre- and post-test with convenience samples), one indicated

---

[1] The relationship of this classification system to that of Campbell and Stanley is obvious.

Design 3a (control group with convenience samples), and one indicated a combination of Designs 4b (pre- and post-test with control group using statistical samples), 5a (a quasi-design using convenience samples), and 6a (a non-standard design using a convenience sample). Designs 2a and 3a are subject to well known deficiencies affecting both internal and external validity, as Campbell and Stanley have shown. Thus only one proposal could be considered to have an adequate design.

b. Means and Samples. Table 2 shows the categories that were used for classifying proposals in terms of data collection techniques used and samples employed. Entries in the cells indicate the number of items so classified. A number of proposals contained multiple technique-sample combinations, of course; hence the total of 43.

Eight proposals gave no indication either of data collection devices or samples; these are shown in Cell 0-0. The remaining 13 proposals yielded 35 classifiable items, and it will be seen that these tend to form several well-defined clusters.

Among the populations from which data might be collected (rows of Table 2), the most popular is, as might be expected, students. Seventeen of the 35 entries are found in this row—about half. The remaining entries are scattered more or less evenly among the remaining rows. Among the techniques for data collection (columns of Table 2), the one most frequently mentioned (12 times) is undefined

TABLE 2
FREQUENCY OF OCCURRENCE OF DATA COLLECTION TECHNIQUES AND SAMPLES

Data to be Collected From:	Data to be Collected by Means of: Can't Tell (0)	Standard Tests (a)	Ad Hoc Tests (b)	Questionnaire (c)	Interviews (d)	Observation (e)	Ratings (f)	Audio or Vis. Record (g)	Case Study (h)	Check List (i)	Judgments	Totals
0-Can't tell	8									1	1	10
1-Teachers				1	1						2	4
2-Students		9	3	1	1	1		1			1	17
3-Administrators											4	4
4-Project Pers.											2	2
5-Parents/Commun.		1			1						1	3
6-Consultants	1					1					1	3
Totals	9	10	3	2	3	2	0	1	0	1	12	43

"judgments," made most often by administrators. The next most frequent column is standardized tests, with 10 entries. No other technique is used frequently, and ratings and case studies are not mentioned at all. The most frequently found single cell is Cell 2-a, standardized tests administered to students.

c. Criteria. Proposals were examined to determine what criteria would be invoked in interpreting data. It is not possible to prepare a reasonably exhaustive taxonomy for criteria since these tend to be highly specific in relation to the nature of the project. In order to make possible at least a nominal classification, however, the following general categories were employed:

1. Numbers. Criteria were classified here if they related to absolute or relative (gain) measures along some obviously numerical continuum, e.g., the number of participants, the number of requests for materials, the amount of use of materials, the number of services offered, the amount of correspondence received, the number of visitors entertained, the number of speeches or publications, etc.

2. Achievement. Criteria were classified here if they related to absolute or relative (gain) measures along some achievement continuum, e.g., knowledge, skills, attitudes, socialization, etc.

3. Amelioration. Criteria were classified here if they

related to some absolute or relative (gain) measure of the amelioration of some undesirable condition, e.g., physical pathology, conditions of disadvantage, inadequate community resources, etc.

4. Acceptance. Criteria were classified here if they related to some absolute or relative (gain) measure along a continuum of acceptance, e.g., involvement, support, commitment, etc.

5. General objectives. Criteria were classified here when they were non-specific but referred to a general "achievement of objectives."

For nine of the 21 proposals it was impossible to make any determination concerning criteria, since none were named. The remaining proposals, some of which had multiple criteria, were classified as shown in Table 3.

TABLE 3

FREQUENCY OF OCCURRENCE OF
VARIOUS CRITERIA

Type of Criterion	f
0 - Can't Tell	9
1 - Numbers	15
2 - Achievement	6
3 - Amelioration	3
4 - Acceptance	3
5 - General	6

The most frequently used class of criteria is thus "numbers," again, as one might expect. The next most frequent specific category is achievement, and usually this is knowledge or skills achievement.

d. Agents. Proposals were examined to determine who it was that was charged with the responsibility for drafting evaluation plans and carrying them out. The data are shown in Table 4.

Outside consultants figure very prominently—12 of the 13 proposals that identified agents included consultant help. In six cases the consultants did it all. Only one system felt sufficiently confident to go it alone without consultant help. State Departments are apparently consulted only infrequently.

TABLE 4

FREQUENCY OF OCCURRENCE OF TYPES OF RESPONSIBLE AGENTS

Type of Agent	f
0 - Can't Tell	8
1 - Local Personnel	1
2 - Consultants	6
3 - (1) and (2) Combined	4
4 - (1) and (2) plus an advisory committee	1
5 - (1) and (2) plus state department	1

Analysis and Interpretation

Several conclusions may immediately be drawn from the above data:

1. Most proposal writers are still uninformed about the emergent meanings and uses of evaluation, as outlined, for example, in the Stufflebeam paper.

2. Most proposal writers seem to be unaware of the criteria that might be applied to an evaluation design to determine whether it is appropriate or inappropriate. As a consequence the typical design is inferior and likely to lead to invalid or unreliable conclusions. It is dubious whether the data from such evaluations will have high credibility, even for the practitioners producing them. Their timeliness and pervasiveness are also in grave doubt.

3. Budgetary considerations are usually lost. Apparently most proposal writers are sufficiently unaware of the budgetary implications of evaluation to be unconcerned about them. As a result no money is available for evaluation in too many cases.

4. Evaluation designs are almost non-existent. Those that are used are typically faulty.

5. Standardized tests and informal judgments are the backbone of most evaluations. Evaluators neglect many useful techniques that are available. Most data are obtained from students.

6. Proposal writers skirt the matter of criteria for

interpreting evaluative data. When mention is made of them it happens too often that only general ideas of "meeting objectives" are presented. When more specific criteria are invoked one is apt to be caught up in a "numbers game" or to be presented with traditional achievement data, usually collected with standardized tests. Many useful even if more subtle criteria are ignored.

7. Consultants are used heavily but with a conspicuous lack of success. It is obvious that outside consultants were used to some degree in the large majority of proposals examined here, but despite this fact, the proposals remain full of loopholes and inadequacies.

Recommendations[2]

It must be clear both as a result of this analysis as well as earlier analyses that the serious deficiencies which are found in the evaluation sections of Title III proposals cannot be ameliorated quickly, and especially not by patch-up, more-of-the-same methods. Fundamental measures must be taken. Strong measures must be directed against the basic lacks that account for and are the root of the persistent failures to do a good job on evaluation.

There seem to me to be three basic lacks: (1) the lack of adequate theory, models, and designs to guide evaluative activity (as evidenced by the fact that even the "expert" consultants do a bad job), (2) the lack of trained personnel (as evidenced by the fact that even the most rudimentary principles of the game are consistently violated), and (3) the lack of appropriate data collection techniques and data processing facilities (as evidenced by the heavy reliance on standardized tests or informal judgments).

---

[2] The recommendations of this section are based upon ideas developed by the present author in lengthy consultation with Daniel Stufflebeam and Stuart Westerlund, now of the Bureau of Research, Office of Education.

Accordingly I recommend that Title III take the leadership in mounting three new agencies designed specifically to deal with these three lacks. These three agencies might be tied together by a common board or by an interlocking directorate. They include:

1. A NATIONAL LABORATORY FOR THE STUDY OF EVALUATION (NLSE).

Such a laboratory should consist of a network of agencies qualified to perform research and development on evaluation problems on a long-range basis. The membership of this laboratory might include such agencies as the Center for Instructional Research and Curriculum Evaluation (CIRCE) at the University of Illinois, the Center for the Study of Evaluation of Instructional Programs (an R&D Center) at UCLA, the Measurement Research Center at the University of Iowa, the Ohio State University Evaluation Center, the National Institute for the Study of Educational Change at Indiana University, and others. Such a network of agencies could deal cooperatively with problems of formulating and testing theory, developing prototype materials and providing leadership for the many agencies that will train the evaluators needed in elementary and secondary education, and developing designs and instruments for the collection and analysis of information by schools, state education departments, OE, and other agencies. The laboratory could also take on shorter range assignments such as developing, testing, and demonstrating prototype training institutes and workshops in evaluation, assist in determining the information requirements of the Title III evaluation program, and develop immediately applicable designs, means, and criteria for ongoing Title III projects.

2. A NATIONAL INFORMATION CENTER FOR EDUCATION (NICE).

The Center would have as its purpose the organization, processing, and reporting of evaluative information. Once the National Laboratory mentioned above has completed the basic research and development for determining information requirements and developing instruments and designs, NICE would take over the responsibility for continuous up-dating in those areas. Additionally, the agency would maintain records of the aims and goals of Title III, the objectives and procedures of program elements within Title III, information resources for Title III projects, information requirements of the major audiences of Title III projects, and in particular, the scope, sequence, timing, locus, focus and criticality of decisions requiring evaluative information. NICE would also be charged with the development of appropriate designs and instruments in advance of their need. The agency would, finally maintain a mechanism for organizing, analyzing, and reporting evaluative information.

3. A NATIONAL GRADUATE SCHOOL FOR EDUCATIONAL EVALUATION (NGSEE).

Existing institutions which might engage in the training of evaluation personnel have not risen to the challenge nor are they likely to under the very limited funding for training programs available under Title IV of ESEA.

Precedents for the establishment of such a school already exist within the military (e.g., the Air War College), the Department of Agriculture, and the State Department. Models for such a school are especially widespread in industry, as for example, the advanced training school of the American Paper Institute at Appleton, Wisconsin, which awards a certificate equivalent to the Ph.D. degree.

The School would be independent of existing institutions and hence would avoid the morass of existing higher degree program requirements within these institutions (e.g., the doctoral language requirements). It could offer programs at a variety of levels beginning immediately on the post-Bachelor level. Training could be available in several blocks to permit persons at a variety of experience and responsibility levels to participate meaningfully. The School could also develop training materials which could be used by its graduates to train personnel in field situations who would be working with, or for, them. By these means the School could vastly multiply its influence.

The School would also afford an opportunity for a more visible career development line for evaluation personnel. After gathering experience, former graduates at lower levels could return for advanced training and then move on to more responsible positions. The school would furnish a highly visible source for evaluation personnel and its placement service could move personnel along to higher level posts as their competence warranted.

The School would tie in with the National Laboratory and the National Information Center for internship and apprenticeship experiences for its students. These latter two agencies would profit from having trained personnel available and could contribute materially to the development of training programs that would better serve emergent needs.

No doubt these three recommendations will appear, on first glance, to be grandiose and impractical. Yet, in my opinion, any measures short of these will at best patch up the existing organizations and programs, and will not produce any long range gains. Drastic measures are needed to cope with drastic problems. The three recommendations seem to me to be the minimal steps required to produce any permanently useful response to the problems of evaluation that now face us.